

What medium for the message?

Can advertising companies, which are so adept at changing consumer behaviour, persuade people to alter their lifestyles to thwart the spread of the AIDS virus?

THE launch of the US government's education campaign to inform the public of the dangers of the virus causing AIDS (acquired immune deficiency syndrome) has been long in coming. More than a year has passed since the United Kingdom and Australia, nations where the magnitude of the epidemic is very much smaller, took similar action. The spread of the virus has continued: a million people in the United States are now believed to be infected and the prospect cannot be discounted that the disease will reach parts of the population that have so far been unaffected. Given these grim circumstances the public has every right to ask whether the government's tardy response has been delivered with all the vigour and forthrightness that will be needed to slow the spread of the AIDS virus and whether it can achieve its goal by the methods chosen.

"America Responds to AIDS", a catchy title and an appropriate one, was designed for the most part by the Centers for Disease Control (CDC), and it has many of the necessary ingredients to raise public consciousness about AIDS. First, there is a slick media campaign by one of the largest advertising agencies in the United States, Ogilvy and Mather. With the skill and guile for which such Madison Avenue agencies are famous, the television and radio pieces lasting no more than a minute present compelling stories of the tragic consequences of being uninformed about AIDS. In addition to segments designed for a general audience, there are also messages specifically aimed at blacks and hispanics, two subsets of the population hit particularly hard by the disease. The advertisements give few specifics about how AIDS is transmitted, or how one might protect oneself from contracting the disease, instead urging people to call a national telephone number to "get the facts". About 45 million copies of a reasonably informative brochure will be printed to accompany the broadcast advertisements. Throughout October, officially proclaimed AIDS prevention and awareness month by President Reagan, there will be daily activities around the country to provide additional information about the disease.

Campaign

But even as the government launched its campaign, there were clear signs of the ambivalence that has characterized the Reagan administration's attitude towards AIDS and the steps that are needed to combat the disease. Just a week before the start of October it was unclear whether the president would take the relatively innocuous step of issuing a proclamation giving White House blessing to the awareness month. The proclamation that did emerge from the White House contains the now familiar moralistic refrain that AIDS can best be prevented by abstention of sexual activity until adulthood, "and then to restrict sex to a monogamous, faithful relationship". This sort of moralistic finger-wagging may accurately reflect the mores of a large segment of the population, but it is not a message that will affect the people most likely to engage in behaviour that will place them at risk. Bible-thumping in an inner-city heroin 'shooting gallery' is a risky occupation that is unlikely to generate a flock of converts.

Any explanation of how AIDS is transmitted must include a

frank discussion of sexual practices. Ogilvy and Mather's broadcast advertisements are hardly explicit; the word condom is breathed in one of the segments, but the television networks that will be called on to air the campaign have already indicated the mere mention of that word is too controversial to be broadcast. For that position, they should be ashamed. But more damaging to the potential impact of the campaign is that there is no money budgeted for purchasing broadcast time. The televised spots fall into the category of 'public service announcements' that broadcasters are expected to air 'in the public interest'. Advertising agencies spend a great deal of time and money determining when to place their advertisements in order to reach their target audiences. The times when key audiences can best be reached are unlikely to be provided free by television networks or local stations around the country. The government knows how much it costs to run an effective advertising campaign. The Department of Defense spends in the neighbourhood of \$150 million on its television-advertising campaign to solicit new recruits. Ogilvy and Mather has a budget of just \$4.6 million: a discrepancy that surely reflects the government's uncertainty over how to proceed with the delicate topic of AIDS. The \$20 million that went unspent from last year's AIDS education budget only adds to that impression.

Education

There is also the question of whether the people seeing the television advertisements and calling the information number are the people most in need of the information. Well educated heterosexual individuals are as a population the least likely to become infected with the virus, and the most likely already to be well informed about the risks of their activities.

Can public-education campaigns change the level of awareness about AIDS? They certainly can. So much is clear from a recent report from HMSO (Her Majesty's Stationery Office) about the impact of the campaign launched in Great Britain. The HMSO report also hints that the campaign changed behaviour, a much more difficult and crucial task to accomplish. But again, there is the very real concern that the people seeing the campaign and changing their behaviour are not the people most crucially in need of altering their behaviour. Among homosexuals canvassed by HMSO, there was a decline in anal sex and number of sexual partners, but the trend towards altered behaviour began before the campaign was launched, suggesting that other factors were responsible for the observed differences. In the United States, homosexual communities for the most part have been exemplary in their response to AIDS. But the virus has had time to work its deadly course in these communities, and it is certain that the death of a friend or lover has more of an impact than a leaflet stuffed in a letter box.

Any AIDS campaign that is to be successful must take into account that people do not like to face unpleasant facts. Denial is a common reaction: people are inclined to believe either that AIDS will not affect them, or that if it does there is nothing they can do to prevent it. In either case, they are unlikely to change their behaviour. Ogilvy and Mather try to address this, emphasizing that AIDS will affect everyone if it gets out of hand.



But an effective campaign will need to be more personal, more clearly targeted and more frank in its approach to change behaviour promoting the spread of AIDS. Although money spent by Ogilvy and Mather will reach a lot of people, there is an urgent need to match these funds with equal or greater sums spent on programmes in urban areas to reach drug users, the likely gateway for the virus to enter the general population. Whatever moral repugnance the administration may feel at dealing with drug abusers, the United States will neglect this group at its peril. □

Flirting with destruction

Mr Amadou M'bow as director-general for a third term is a recipe to put UNESCO out of business.

INSTITUTIONS which the gods would destroy they first make irresolute. That seems to be the principle on which UNESCO, the United Nations Education and Scientific and Cultural Organization) is now courting extinction. The circumstances are these. The director general for the past ten years, Mr Amadou M'bow, is due to retire at the end of the year. For several months, UNESCO has been seeking a replacement. Perhaps to its executive committee's surprise, there have been several candidates, many of them well qualified for the job. One, for example, is professor Abdus Salem, awarded a Nobel prize for his contribution to the synthesis of electromagnetic and weak nuclear interactions and now director of the National Institute for Theoretical Physics at Trieste. With such candidates willing to shoulder the herculean task of putting UNESCO back on a solid foundation, it is beyond belief that UNESCO's managers should have such difficulty in making a suitable choice. yet, now, the word is that the executive committee cannot make up its mind, and that M'bow may be given a third term.

That will be a recipe for UNESCO's end. M'bow is a controversial character who is largely and personally responsible for the defection of three of UNESCO's members in the past two years — the United States, the United Kingdom and Singapore, in that order. The reasons are several and complicated. M'bow's first term in office was marked by a long dispute about the rights and wrongs of the infamous "world information order", designed as a scheme for improving the flow of information within the shrinking world that would also have had the effect of enabling governments to regulate the flow of information about themselves. The scheme was a legitimate topic for UNESCO's interest, but it was foolish of M'bow to have persisted with it when it had become clear that the scheme offended against western principles of what constitutes a free press. (It would be interesting to know where the Soviet Union, one of M'bow's supporters in that argument, now stands, with glasnost in the air.)

This is not the only complaint against M'bow during his two terms of office. The administration of his sprawling organization, never good, is at a lower ebb than ever. The proportion of UNESCO's shrunken budget spent on its own staff is greater even than it was. There is probably also some truth in the complaints that M'bow's way of doing business has led him to keep his staff enthralled by unjust exploitation of the need to employ people on short-term contracts. Arguments on these questions will persist, but they have no bearing on whether M'bow should be reappointed. The simple truth is that the departed members of UNESCO will not return while he remains in office and that his continued presence will probably drive others away. The African governments that support him may consider that a scandal, but it is the reality to which they have to adapt if they wish to keep UNESCO in being. It would be especially absurd that M'bow should be reappointed when there are alternative candidates whose reputation has not been tarnished by the troubles of the past few years. □

The art of the soluble

Peter Medawar should be remembered as much as a writer as for his contributions to immunology

FOLLOWING his death last week, at the age of 72, much will rightly be made of Sir Peter Medawar's fundamental contributions to the science of immunology, of which the most important was the demonstration that the ability of the immune system to distinguish 'self' from 'non-self' is not inherent but is learnt during the system's development as a result of exposure to 'self' molecules. But for both those who intend to, and those that already, toil at the bench or in scientific committees, Peter Medawar's words of wisdom and advice are of equal importance. Fortunately, they are preserved in his extensive writings. Even more fortunately, Medawar was unequalled as a writer among scientists.

It was *The Art of the Soluble*, published 20 years ago, that brought Medawar's skills as an essayist to the scientific community and beyond. The title of the book followed the lead of Bismarck and Cavour, who had explained the art of politics as "the art of the possible". Later, Medawar was to complain that his description of the art of the research as that of the art of the soluble had been almost wilfully misunderstood by some people to mean that he had advocated the study of easy problems yielding quick solutions, adding, in typical style, "unlike my critics, who were studying problems of which the main attraction (to them) was that they could not be solved".

Medawar's point was that science proceeds best when experiments test well-formulated ideas. "Clearly", he wrote, "a hypothesis so permissive as to accommodate any phenomenon tells us precisely nothing; the more phenomena it prohibits, the more informative it is." His extensive writing on the scientific process — of how "no new truth will declare itself from inside a heap of facts" — owed much to Karl Popper. Medawar was the first to acknowledge this; but he was also the first to enable those scientists with the nous to realize that process matters, but not the stomach for philosophers' academic writings, to appreciate what Popper has to say in a context that strikes a chord.

In his *Advice to a Young Scientist* (Harper and Row, 1979), Medawar returned to this theme: "It is a common failing — and one that I have suffered from — to fall in love with a hypothesis and to be unwilling to take no for an answer. A love affair with a pet hypothesis can waste years of precious time. There is very often no finally decisive yes, though quite often there can be a decisive no . . . a scientist does not very often speak with complete confidence of proof. The more experienced he is, the less likely he is to do so". Such views were naturally linked to another of Medawar's favourite themes — the ways in which scientific papers obscure the scientific process. The conventions and constraints of the primary paper, he emphasized, allow little if any mention of the false trails, errors, serendipity and failed hypotheses that inevitably litter the pathway along which any successful scientist must tread. *Is the Scientific Paper a Fraud?* asks one of his most famous essays.

In his advice to the (potential) young scientist, Medawar repeats his favourite intelligence test: "To many eyes, some of the figures (particularly the holy ones) of El Greco's paintings seem unnaturally tall and thin. An ophthalmologist who shall be nameless surmised that they were drawn so because El Greco suffered a defect of vision that made him see people that way, and as he saw them, so he would necessarily draw them. Can such an interpretation be valid?" As a rider, and one that captures much of the quality of the man and his writing, Medawar adds, "Anyone who can see *instantly* that this explanation is nonsense and is nonsense for philosophic rather than aesthetic reasons is undoubtedly bright. On the other hand, anyone who still can't see it is nonsense even when its nonsensicality is explained must be rather dull." Rereading Medawar periodically will keep dullness at bay. □

Ozone hole deeper than ever

- New antarctic data prove chemical cause
- Global significance not yet clear

Washington

New data from the US Antarctic Ozone Project, released in preliminary form last week, will no doubt help to settle the scientific debate over the chemical and meteorological causes of the ozone hole that recurs over the South Pole each austral spring. But whether they will calm the political debate over what should be done is a different matter. In brief, the latest results demonstrate an undoubted chemical cause in the destruction of ozone by atmospheric chlorine, but also point to special climatic conditions as the reason why depletion occurs so severely in the antarctic and so little elsewhere. This leaves room for environmentalists to demand action and for chemical manufacturers to emphasize the lack of imminent global danger.

The project, managed by the National Aeronautics and Space Administration (NASA) with help from other US agencies, foreign countries and the US Chemical Manufacturers Association (CMA), combined airborne and ground-based investigations (see *Nature* 328, 463; 1987). At a press conference on 30 September, only days after the end of experimental flights, Robert Watson of NASA and Dan Albritton of the National Oceanics and Atmospheric Administration (NOAA) described some of their preliminary scientific findings ahead of publication, in recognition of intense public and political interest.

The concentration of ozone at 70 and 80° S was about 15 per cent lower this year than in 1985, the greatest depletion recorded previously. This is consistent with a suspected trend of increasing ozone loss, modified by a two-year cycle. Two particularly telling details stood out. As ozone destruction proceeded during September, the atmospheric concentration of chlorine monoxide showed a proportional increase; this clearly implicates chlorine chemistry. But the decrease in ozone levels did not proceed smoothly either geographically or with time. On 5 September, a region over the Weddell Sea showed a fall in ozone level of more than 10 per cent in a 24-hour interval; such an

anomaly must, said Watson, reflect local weather conditions, not chemistry. In an official statement, the CMA agreed that the antarctic ozone depletion was a consequence of "unique meteorology and chlorine chemistry".

Instruments aboard a DC-8 flying at 10–12 km in altitude, and an ER-2 (a modified U2 spy plane) at up to 18 km measured concentrations of ozone, chlorine and nitrogen compounds as well as the compositions of stratospheric ice crystals. Taken together, these measurements confirm the peculiar nature of the upper antarctic air, which is not only dehydrated, because water is locked up in ice crystals, but also denitrified (nitrogen oxides are adsorbed onto the ice crystals as nitric acid, and precipitated out of the atmosphere). In normal air, nitrogen oxides are able to lock up chlorine in the form of nitrates, which then cannot destroy ozone. In the denitrified antarctic air, chlorine does other things.

The exact route by which chlorine destroys ozone is still somewhat mysterious. Watson reported that measurements of bromine compounds yielded concentrations of only a few parts per trillion, which inclined him to doubt the importance of a chemical mechanism in which bromine monoxide acts as a catalyst in the destruction of ozone. But Michael McElroy of Harvard University disagrees, saying that even at the reported low concentrations the bromine mechanism would still be more effective than the comparable series of reactions involving chlorine alone. Mike Kurylo, a joint programme manager at NASA, preferred not to favour one mechanism over another, but said that laboratory studies of the rates and yields of the important reactions should, in conjunction with the complete atmospheric data, pin down the chemistry of ozone loss.

Although Watson and Albritton adamantly refused to go beyond their preliminary data, and said it would be unwise to think about global consequences or possible international action before the scientific results are analysed and published, others are not so reticent.

McElroy, convinced that bromine is critical, is concerned that even a complete cessation of chlorofluorocarbon (CFC) production would leave a reservoir of chlorine in the atmosphere for a century or more, and that bromine emissions must also be halted to prevent continuing erosion of the ozone layer. At present, the two chief sources of atmospheric bromine are methyl bromide, half of which is industrial and half from marine organisms, and halons, chemicals used in nondestructive fire-extinguishers, particularly in the electronics industry. McElroy says that these sources have been increasing at more than 10 per cent per year recently, and may indeed be behind the downwards trend in ozone levels at the south pole.

Irving Mintzer of the World Resources Institute is more optimistic. The newly agreed international protocol on the protection of the ozone layer (see *Nature* 329, 277; 1987) provides for a quadrennial scientific review, as well as for emergency reviews by request of a few of the participating nations. New scientific data will thus be automatically accommodated. And Mintzer sees the participation of the CMA in this venture as recognition that there is now 'market pressure' on industry to develop safer CFCs which do not persist for so long in the atmosphere.

Direct global implications of the ozone hole are hard to fathom. The new data seem to imply that the 'chemically perturbed' region corresponds more closely to the area of the antarctic continent than to the area of meteorological oddity, and no scientist is claiming that there is any danger of geographical spreading of the hole. For Mintzer, the significance of the ozone hole is that it has drawn attention to the spectacular way in which human action can alter a piece of the Earth's environment, and at a rate beyond human ability instantly to reverse. David Lindley

GenBank takeover

INTELLIGENTICS Inc., of Mountain View, California, will take over from Bolt Beranek and Newman Inc. as the administrator of GenBank, the government-sponsored database of genetic sequence information. The Department of Health Services last week announced the five-year, \$17.2-million contract. The National Institute of General Medical Sciences is the main funding agency for GenBank, and manager of the project. Intelligentics will collaborate with Los Alamos National Laboratory to implement the database. The new contractors are hoping to encourage researchers to submit sequence data to GenBank in computer-readable formats, bypassing slower manual data entry. J.P.



NASA's DC-8-72 flying laboratory.

Space science celebrated on Sputnik's thirtieth birthday

Moscow

THE international space research community last week celebrated the 30th anniversary of the first Soviet Sputnik, launched on 4 October 1957, with a symposium on the future of space research destined, from the outset, to turn into a gigantic party. By the end, most participants were convinced that the next step forward is an international collaboration for the exploration of Mars, but official government cheque-books were conspicuous by their absence.

Last week's event was the brain-child of Academician Roal Z. Sagdeev, the plasma physicist turned manager of the Soviet civil space research programme, and now director of the Space Research Institute of the Soviet Academy of Sciences. Sagdeev was described on Sunday by a member of the US delegation as the man whose enthusiasm for international collaboration now promises to set its stamp on the pattern of research in the United States and elsewhere.

More than 250 people from outside the Soviet Union, together with an equal number of Soviet scientists, were assembled for the celebration. A genial Soviet cosmonaut explained that the gathering included 41 astronauts and cosmonauts (the Soviet equivalent), the largest gathering of such people "in the history of the business".

The tenor of the proceedings was remarkable for its amity. Soviet speakers went out of their way to congratulate the United States for its achievements during the past 30 years, the "brilliant success" of the Apollo programme included. One Soviet speaker even drew attention to the hazards of the space business, emphasizing the difficulties of running manned spaceflights on a tight schedule.

The formal proceedings finished with a conversation between two orbiting members of the Soviet squad, three ground-based astronauts (one from the United States), Sagdeev, Academician G.I. Marchuk, president of the Soviet Academy and Mrs Carl Sagan (whose husband was taken ill after delivering an upbeat address to the symposium). The

Japanese finance: correction

AN article in *Nature* of 17 September (329, 188; 1987) stated in regard to the reorganization of the Astronomical Observatory of Tokyo University into a national research institute that "funds for reorganization were requested this fiscal year but were turned down by the Ministry of Finance". In fact, a small budget for "preparation for reorganization" was allotted in this fiscal year. □

conversation was a telling reminder of the difficulty of avoiding banality on these occasions.

The meeting's formal decisions were few, but none the less interesting on that account. There is to be a study of the potential usefulness of satellites as a means of relaying data of urgent significance to developing countries — epidemiological data about AIDS (acquired immune deficiency syndrome) and information about natural disasters, for example. Two groups, based in Boston and Moscow, will try to devise a workable scheme by the end of the year.

The United States plans to hold another conference on future directions in space research in the middle of next year. The V International Association of Space Artists

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Sputnik 1 — the first official photograph of the 30-year-old Earth satellite released from Moscow on 9 October 1957.

(whose members exhibited their largely representational works at the celebrations) will arrange a travelling exhibition in the months ahead. The conference adopted a resolution (read by Ms Susan Einsenhower, whose grandfather was US president at the time of the first Sputnik) urging collaboration in the peaceful uses of outer space, particularly during the international space year planned for 1992.

Enthusiasm for the exploration of Mars has been largely stimulated by Soviet plans for the years ahead. The rocket for the flight to the martian satellite Phobos will be launched next year, but Soviet participants at last week's conference gave a detailed account of the missions they plan to send to Mars beginning in 1992. The plan is that a series of rocket flights should recover information about the surface of the planet by remote sensing and that there should be a series of landings on the surface of the planet by various roving instruments with the intention that samples should eventually be recovered. The most intriguing candidate design for the surface vehicle is one consisting of a gas balloon and a hot-air balloon linked together so as to drift above the surface during the martian daylight. The design is being developed jointly by Soviet and French groups.

The symposium also provided a valuable summary of some recent astrophysical results. The discovery of a curious gravitational lens in which a pair of images is superimposed on a diffuse ring of radio luminosity was reported from the Very Large Array telescope in the United States, and various luminaries such as Ya. B. Zeldovich and Andrei Sakharov provided a perceptive account of the state of cosmology. The principle seems to be that ambitions should be unconstrained.

John Maddox

IVF rules under pressure to change

London

PRESSURE for rapid legislation on *in vitro* fertilization and embryo research is mounting in Britain following the flouting by a London clinic of voluntary guidelines. The privately funded Humana Hospital Wellington ignored guidelines imposed by the Voluntary Licensing Authority (VLA) by transferring more than four eggs at a time. The hospital has been struck off the VLA's list of approved clinics but remains free to continue to practise because the VLA has no statutory powers.

The VLA announced the new guidelines in May (see *Nature* 327, 92; 1987) and gave clinics until September to bring their practices into line. The Wellington was the only one of 30 *in vitro* fertilization clinics in Britain that refused to provide the necessary assurances.

There are indications that the VLA

might be prepared to reconsider one of its new guidelines (also announced in May) forbidding the donation of eggs by the relatives of infertile patients, in line with its principle that the anonymity of the donor is paramount in the interest of the child. The VLA now acknowledges that some clinics feel that this rule should not be inflexible. The question will be raised at the clinics' annual meeting next month.

The VLA was set up two years ago by the Royal College of Obstetricians and Gynaecologists and the Medical Research Council following the completion of the Warnock report, commissioned by the government, on human fertilization and embryology, which recommended a statutory licensing authority. A long-awaited government white paper (policy statement), based on the Warnock report, is due in late November. Simon Hadlington

Science for Britain set to back a "free world perspective"

London

SIR Ian MacGregor, former chairman of Britain's National Coal Board, last week announced the formation of a new pressure group, Science for Britain. The new group, of which MacGregor is president, has the support of the Watt committee, which represents 62 institutions with an interest in energy, and the British Science and Technology Trust, a charity aimed at encouraging more young people to embark on careers in science.

Another backer is the recently formed Advanced Energy Research Institute (AERI), a private organization founded by MacGregor and Canadian businessmen Leonard Holihan, who intends to identify and secure financial backing for potentially exploitable ideas and inventions "from a free world perspective".

According to the leaflet announcing Science for Britain, innovators and their concepts frequently leave Britain, impoverishing its ability to compete internationally "to protect and promote our freedom". The group also calls for encouragement of international participation in "the strategic vastness of space development". One of its stated aims is to "emphasize Britain's role in spearheading technological advancement in the western democracies and to promote awareness of the direct link between scientific progress, jobs, prosperity, peace and freedom". The leaflet invites voluntary donations to the campaign.

AERI is currently devoting most of its attention to high-temperature superconductivity. Indeed, for its launch, at the British Science Museum on 15 July, Holihan attracted the interest of researchers and the media when he promised to announce a significant breakthrough in superconductivity. The 'high-current density' ceramic cylinder he displayed, manufactured by a small four-man company AERI has taken under its wing, failed to impress the academics.

Holihan and MacGregor met a few years ago and discovered they shared an interest in "catalysing future technological developments". They formed, in 1983, the Aeronautical Research Institute (ARI), which was particularly interested in "small vertical take-off and landing conceptions" (more popularly known as flying saucers). AERI is a successor to ARI, with wider aims.

According to Holihan, AERI identifies potentially exciting projects through its network of contacts, and then tries to interest potential backers. It is looking to accelerate technological developments from a "free-world perspective". To explain that term, Holihan says: "There is

the free world and there is the Soviet bloc. Science does not happen in a vacuum devoid from politics or totalitarianism. We should all be tuned in to what the Soviets are working on as far as technology to avoid a technological run-in. The free world has to be astute as to what the strategies and priorities are in regards to scientific technology to protect and preserve our freedom. It is not just science for science's sake".

In 1981, Holihan and two Conservative politicians set up an organization called 'Coalition for Peace through Security', a right-wing pressure group opposed to the growing support for the unilateral disarmament lobby in Britain.

Shortly afterwards he directed a 'private think tank on strategic affairs'.



MacGregor announces a new pressure group avowed to "emphasize Britain's role in spearheading technological advancement".

AERI is in the process of establishing a consultative council. The only name Holihan is prepared to release at this stage is that of Sir Kenneth Corfield, director of Midland Bank, Britoil and Octagon Group. Holihan says "a few more" people have agreed to join the consultative council, but their names will not be announced until the full establishment of 12 has been reached.

Holihan says AERI is looking seriously at "a dozen or so" projects, and although no funding has yet been secured, several potential backers, whom Holihan declines to name, are "interested". He says: "Sir Ian and I are passionately interested in science and technology and its practical applications to help people. It's good for business and it's good for strategic reasons." Asked if AERI is interested in military applications of technology, Holihan said: "We are not particularly concerned with weaponry but obviously anything has some applications towards that." Holihan says that AERI is "absolutely unpolitical. I am not a political person. This work is trans-political".

Simon Hadlington

French suspect information on radiation levels

Paris

ACCORDING to a survey carried out for the French Industry ministry, only 35 per cent of those interviewed believed the government would be ready to protect the safety of the population in the event of a major nuclear accident, such as occurred at Chernobyl last year.

Almost half (49 per cent) of the sample were "very" or "quite" worried about the likelihood of such an accident in France. This represents a marked change in the climate of public opinion in France since 1985, when only 33 per cent of those questioned said they were concerned.

For France, which is second only to the United States in its investment in nuclear energy — with few naturally occurring alternative energy sources — public confidence has been a significant factor enabling successive governments to continue to build reactors when all of France's European neighbours slowed or halted construction because of public pressure.

All official statements about radiation levels are routed through the government 'mouthpiece', the SCPRI (Service Central de Protection Contre les Radiations Ionisantes). Following the Chernobyl accident, it was the SCPRI who initially denied that the radioactive cloud had passed over France and, subsequently, published local measurements of radiation levels in soil which were 80 per cent lower than those recorded by Electricité de France, who operate all nuclear reactors in France.

Doubts over the accuracy of figures released by SCPRI have led to the creation of an independent body, CRIIRAD, which periodically monitors levels of radiation in foodstuffs. With the arrival of autumn, when French families traditionally go mushrooming, CRIIRAD has recorded high levels of caesium 137 and caesium 134 (up to 24,000 becquerels per kilogram) in certain edible fungi collected in the east of France. But as records of radiation levels in these fungi before the Chernobyl accident have not been kept, it is impossible to interpret the new measurements.

Another group, composed of nuclear scientists and calling itself the Groupe-ment de Scientifiques pour L'information sur L'énergie Nucléaire (GSIEN), has been set up to inform the public on nuclear energy problems. One of its aims is to lobby the International Commission on Radiological Protection for a lowering of recommended 'safe' dose levels of radiation to take account of recent research (see *Nature* 329, 96; 1987). Peter Coles

White House science advisor takes the reins after slow start

Washington

ONE year into his job as President Reagan's science advisor and director of the Office of Science and Technology Policy, William R. Graham seems ready to take a somewhat higher profile in national and international science policy-making. Graham is a softly spoken man who has so far shunned the limelight that was much loved by some of his illustrious predecessors. His first important public appearance was not until the summer when he shared the stage with Reagan and the Secretaries of Defense, Energy and State during the opening speeches at the government-sponsored meeting on the commercial applications of superconductors (see *Nature* 328, 469; 1987).

Reagan nominated Graham to replace George A. Keyworth II as his science advisor in June 1986, six months after Keyworth left the post. Inexplicably, the White House neglected to send Graham's nomination to the Senate for months, and it was not until 1 October that the Senate confirmed Graham's nomination. But now it seems that Graham has access to White House inner circles, including the president whom he describes as having a "strong interest in science and technology" which at times "stretches" Graham's capacity to keep up with his questions. On 1 October, for example,



William Graham — opening up

when Graham, trained in physics and electronics, quickly had to become an Earth scientist. The president wanted an immediate briefing on the implications of that morning's earthquake in Los Angeles (see p.479) for other geological faults in the region.

Judging by Graham's activity in two of this summer's key domestic issues — how the spread of AIDS (acquired immune deficiency syndrome) should be contained, and how best to stimulate research on the new superconductors — he seems to see his role, at the top of the pyramid, as one of ensuring efficient inter-departmental coordination. On the international front he is more of an innovator.

Graham sees science and technology, particularly basic science, emerging as

important new factors in international relations. But unlike other key areas of diplomatic activity no international body exists to look after the interests of basic research. Graham looks towards "establishing an international framework for basic science" with the first steps being taken at the OECD (Organization for Economic Cooperation and Development) ministerial meeting in Paris at the end of this month. He looks for support for the notion that basic scientific research is a shared responsibility for all industrialized countries. Guaranteed access to basic research laboratories, regardless of the institutional structures in which they are found, is one component of that responsibility. Another, difficult to realize, will be to find a common view on intellectual property rights so that they do not become a barrier to free exchange of research.

Is Graham thinking of Japan as a

country that has evaded its responsibilities of support for basic research? He says that "Japan shouldn't be taken as a role model for other countries". Rather than taking their basic research from elsewhere, developing countries should build basic research infrastructure alongside industrial development. The special problems that the United States has with Japan will be taken up this month in formal negotiations over the US-Japan science cooperation agreement.

To match Japan's vigour in creating new products, Graham agrees with the thrust of the National Science Foundation's initiatives on giving industry better access to the universities' basic research. "US industry", he says, "must become more aggressive in assimilating basic research" and make room for it "in long-range plans". The role of government he sees as a "catalyst" in a two-way flow between universities and industry. But he does not endorse any one unique solution; "the United States is big enough to tolerate quite a bit of diversity".

Alun Anderson & Joseph Palca

California decides to speed up testing of new AIDS drugs

San Francisco

CONCERN that new drugs for the treatment of AIDS (acquired immune deficiency syndrome) are not reaching the market place quickly enough has led the state of California to declare plans to establish its own drug-testing programme, independent of the federal Food and Drug Administration (FDA). Legislation for the programme, passed by the state government last week, was welcomed by AIDS sufferers' organizations which have long seen the FDA as lacking open-mindedness and a sense of urgency when considering new drugs. But critics contend that the real problem is that there are simply not enough good candidate drugs.

A panel should be at work by the end of the year reviewing applications for in-state human testing of drugs made in California, and possibly licensing the in-state use of drugs that show effectiveness in human trials.

Paul Volberding, an AIDS drug researcher at the University of California at San Francisco (UCSF), and a member of the drug-review panel, said the law will add an unnecessary layer of bureaucracy and he worries that the panel lacks the necessary expertise to evaluate preliminary animal testing of new drugs, and may give approval for human trial to drugs that are unsafe.

Volberding said the FDA has streamlined the approval process for AIDS drugs by working with drug manufacturers and testers. In the case of AZT, now licensed

under the name of Retrovir, permission to begin human studies was granted within a few days of receipt of the application.

Although supporters of the law have faith in the expertise of the Californian panel, they hope it will be prepared to consider new drugs and drugs such as ribovirin, dismissed by the FDA after two trials, one of which gave a positive result, but was shown to be flawed.

Hundreds of AIDS patients are individually obtaining ribovirin in Mexico, said California state assemblyman Bill Filante, author of the bill, but no data can be gathered on their progress. Filante wants to see some of these people brought into a California study, so they can be medically supervised and the data used for evaluation of the drug's effectiveness.

Some critics of the law worry that the exclusive testing of drugs in California will draw AIDS patients to the state, but Filante points out that California is already a magnet because of its highly developed system for AIDS care.

Marcus Conant, an AIDS researcher at UCSF, noted that the law may bring industry to the state, if companies choose to move their drug-developing operations to take advantage of California's testing.

A source of funding for the drug testing has not yet been assured, although \$500,000 has been appropriated for start-up costs. Governor George Deukmejian recently vetoed a bill that would have provided tax credits to those who contributed funds to the programme. **Marcia Barinaga**

Stanford animal-research facility expansion hit by local protests

San Francisco

OPPOSITION to research in genetic engineering from environmental and animal rights groups seems likely to slow Stanford University's plans for expansion of its biological sciences facilities. The ground-breaking ceremony for the new \$33-million building, scheduled for next month, is now under threat following a local group's request for an environmental impact report (EIR) to assess the hazards of recombinant DNA research to be carried out in the building.

The setback comes on the heels of a successful attempt by the local humane society to block construction of the second phase of a \$29-million research animal facility. Construction of the animal facility was scheduled to begin last June. In May, the Palo Alto Humane Society filed an appeal challenging the university's use permit and requested an EIR.

The humane society raised the issue of possible hazardous emissions from the facility, but most of their concerns focus on animal welfare. Stanford argued that the construction of the 'state of the art' animal facility is evidence of their commitment to animal welfare, and provided assurance that the building will meet all current standards for laboratory animal care. Because of the degree of public controversy, the county supervisors nevertheless called for an EIR. Construction has been set back almost a year at a cost to the university of some \$1.5 million in delays and expenses for the report.

The biological sciences building, part of the Near West Campus project, (see *Nature*, 329, 281; 1987) has been on the drawing board for nearly 4 years. On 13 August a county planning committee waived the need for an EIR. The Silicon Valley Toxics Coalition filed an appeal on 28 August to reverse the decision. At a meeting on 1 October, the county planning commission heard both sides of the argument, and planned to vote on 5 November. The commission appears to favour Stanford, but the toxics coalition has vowed to appeal their case in the courts if necessary.

The 6-year-old toxics coalition has never before opposed recombinant DNA research, but was alerted to the issue by the Laurel Heights neighbourhood association, which is fighting expansion by the University of California, San Francisco (see *Nature* 328, 656; 1987). The coalition wants assurances that Stanford will strictly enforce laboratory safety regulations, and that neither harmful levels of radioactivity nor viable recombinant organisms will leave the building.

Stanford claims a good record of com-

pliance with safety regulations, and says no hazardous research will go on in either the new or the existing biology buildings.

Stanford officials will decide next week whether to begin an EIR immediately or to continue to defend their position. An EIR could take up to a year and cost

\$500,000, but a court fight could increase time and cost significantly more.

Stanford spokesman Larry Horton points out that an EIR will provide no new information about the safety of recombinant DNA. Professor Robert Simoni, chairman of the biology department building committee, fears that even a perfect EIR, showing total compliance with safety rules, will be challenged by the groups in their effort to tighten local regulation of facilities.

Marcia Barinaga

Cable companies tied up in knots in international Pacific network

Tokyo

Two rival international telecommunications companies, International Digital Communications (IDC) and International Telecom Japan (ITJ), have filed separate applications for operating licenses with Japan's Ministry of Post and Telecommunications. Meanwhile, Kokusai Denshin Denwa (KDD), which currently monopolizes Japan's international telecom market, has revealed ambitious plans to lay a network of trans-Pacific optical fibre cables to compete with a cable to be laid by one of the new companies.

When Britain's Cable and Wireless company announced plans last year to form IDC in collaboration with Japanese and US companies, there was considerable opposition to their plan in government and business circles. The Ministry of Post and Telecommunications objected to the British firm having a major say in the management of the company and tried to force a merger with the all-Japanese consortium

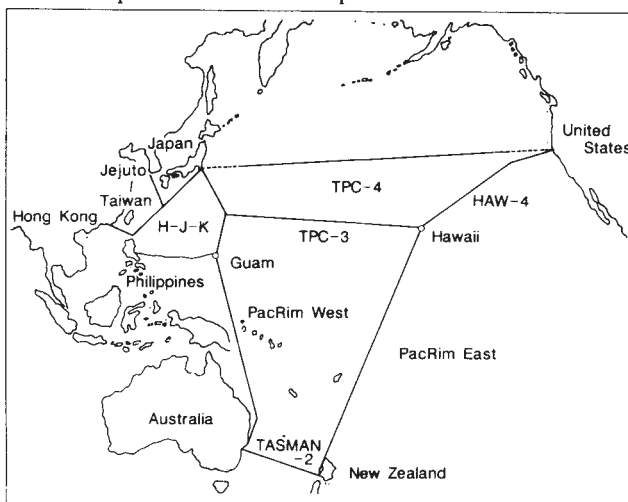
ITJ. The ministry and KDD also argued that there is no need for the trans-Pacific cable which Cable and Wireless intends to lay between Japan and Alaska.

But the exclusion of Cable and Wireless threatened to ignite another major trade dispute with Britain and the United States at a time when the US Congress is compiling an omnibus trade bill, and recently Prime Minister Yasuhiro Nakasone and the ministry agreed to accept a separate license application from IDC and process it in a "fair and transparent manner" (see *Nature* 328, 563; 1987).

IDC and ITJ plan to begin services in 1989 and offer rates 20-30 per cent lower than KDD. Both will use Intelstat satellites, but, whereas ITJ will plug into the

trans-Pacific cable network to be laid by KDD, American Telephone and Telegraph (AT&T) and telecommunications companies from several other countries, IDC will lay its own North Pacific cable in 1989-90 that will short-circuit KDD and AT&T's trans-Pacific cable (TPC-3), due to be completed next year, by 5,000 km.

In response, KDD, which initially argued that Cable and Wireless's cable would result in overcapacity, has announced plans with AT&T to lay



Optical fibre cables to be laid by KDD, AT&T and companies from several other countries. In addition to these cables, Britain's Cable and Wireless and IDC plan a cable linking Japan and Alaska.

another trans-Pacific cable (TPC-4) direct to the United States that will be operational in 1994. It has also agreed with Australian and New Zealand companies to lay a gigantic 16,000-km loop of cables linking Guam, Australia, New Zealand and Hawaii to the TPC-3 cable.

The KDD/AT&T trans-Pacific cables will cost more than \$1,000 million, excluding TPC-4, the cost of which has yet to be announced, compared with \$400 million for Cable and Wireless's 8,000-km cable.

The Ministry of Post and Telecommunications fears that the large number of cables will lead to overcapacity and excessive competition. But Sir Eric Sharp, chairman of Cable and Wireless, says they will create more demand. David Swinbanks

Vision of industrial and rural harmony comes closer to reality

Tokyo

WHAT is a technopolis? When Japan's Ministry of International Trade and Industry (MITI) dreamed up the concept several years ago to revitalize the nation's remoter regions and provide new drives for industry, local governments swamped the ministry with applications for technopolis status. No technopolis is yet complete, but one in the southern island of Kyushu, which got off to a head start, indicates what the future holds.

The MITI plan, which became law in 1983, envisages non-polluting high-technology industries set in green fields in coexistence with farming and fishing communities, a 'mother city' blessed with universities and research institutes that would feed industry with new ideas through an information network, while high-value-added goods are shipped out through a nearby airport.

Twenty-two local authorities have won technopolis status, which allows them to offer tax incentives and low-interest loans to attract new industry. The target date for completion of infrastructure is 1990.

The technopolis under development in Oita Prefecture on the north-east coast of Kyushu got under way ahead of its rivals. The mountainous and forested prefecture is a major tangerine-production area and is also a mecca of hot-spring resorts, but until the 1970s it was suffering from a steady decline in population.

In 1979, the prefecture's newly elected governor, Morihiro Hiramatsu, a former MITI official, launched the airport-based industrial project. Oita's offshore airport was extended to accommodate cargo jets from Tokyo and Osaka, and high-tech industries began flocking to the area.

Now the peninsula boasts more than 30 high-tech companies within 50 km of the airport. Many of them are producers of integrated circuits, such as Toshiba, Matsushita Electric and Texas Instruments. Kyushu produces 40 per cent of Japan's integrated circuits, or 10 per cent of world production, and a large proportion comes from this region, nicknamed 'silicon peninsula'.

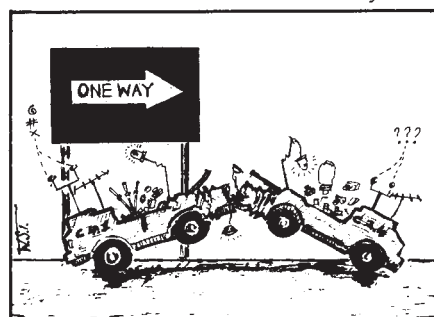
Parallel with the airport project, Hiramatsu launched a 'one-village one-product' campaign which has helped revitalize the prefecture's farming and fishing communities. No prefectural funds were provided, only "moral and technical support". Each village was encouraged to produce its own characteristic product, such as shitake mushrooms, prawns and wine made from kiwi fruit. Although kiwi wine has yet to catch on, much of the prefecture's produce is shipped out along with the integrated circuits in cargo planes

destined for Tokyo.

A stone's throw from the airport, Canon's plant, which manufactures "auto-boy" cameras, provides a glimpse of technopolitan life on the factory floor. The plant is 50 per cent automated. Nine unmanned vehicles trundle around delivering parts to assembly robots and picking up finished goods. The vehicles have flashing lights, bumpers and play Scottish and Irish tunes to avoid knocking over unsuspecting technopolitans.

Most of these ideas come from the workers themselves — they are asked to submit two suggestions for improvements every day. Those who hit on a bright idea can win up to ¥50,000 (more than \$300). Since the factory opened in 1982, output has increased sixfold.

But an essential element in MITI's vision of a technopolis is lacking. Nearly all the high-tech industries are production facilities; the research and development units remain rooted in the Tokyo area. Hiramatsu admits that progress on this front has been slower. But he says that



some joint projects, for example, to design new integrated circuits, have been set up between Oita University and industry.

In 1985, the technopolis's 'mother cities' of Oita and Beppu won designation from the Ministry of Post and Telecommunications as 'teletopias', which thereby entitles them to priority treatment in the establishment of a regional information network system (see *Nature* 305, 36; 1983). And an optical fibre cable is being laid between the two cities. By fiscal 1988, under MITI's 'new media community project' the information network will be extended to all parts of the technopolis via, for example, community antenna television. But it is yet to be seen what type of information will flow in the network.

So what is the secret of Hiramatsu's success in attracting new industries and government projects to his prefecture? The governor says that the contacts he established in his former post as director of the electronics policy division of MITI have been "very important". Furthermore, the presidents of the companies "appreciated his hospitality". **David Swinbanks**

French pirates challenge law

Paris

THE prosecution of a lecturer in computer science at the Paul Sabatier University of Technology, Toulouse, for illegal copying of computer programs, may lead to a change in the copyright laws. The accused, Claude Chrismont, says that the programs were copied for teaching purposes only and that the high price of commercial software, compared with relatively modest teaching budgets, makes such piracy not only necessary, but commonplace. The case has now taken on a national dimension, with hundreds of fellow computer-science lecturers publicly admitting similar offences.

Chrismont was charged following the prosecution of two of his students, in April this year, for selling copies of FF7,000 (\$1,150) software for FF700. The students said it was Chrismont who had provided them with the original copies, as part of their course material.

In a statement published in the French newspaper *La Dépêche*, Chrismont's colleagues, and other senior academics from institutes in Toulouse, admitted breaking the same 1985 copyright law, which permits the purchaser of software to make one back-up copy only. The signatories have not only challenged the authorities to take action against them too, but have also written to computer-science departments throughout France, calling for support. Within two days, more than 300 'confessions' and requests for prosecution had been received.

The affair has now reached the Minister for research and higher education, Jacques Valade, who was visiting Toulouse last week, but it will not be easy to amend the copyright act. The Agence pour la Protection des Programmes (APP), who brought the initial case against the two students in April, is adamant that no change in the law is necessary. The president of APP said that "piracy is costing the software industry FF1,200 million (\$197 million) a year, on a turnover of FF10,200 million. The contribution of copying for teaching purposes is difficult to gauge, but in any case is unjustified".

The APP has suggested lecturers use special versions of software which have been 'spiked' against commercial applications — such as accounts packages handling only 20 transactions. University lecturers, however, argue that they have a responsibility to provide students with realistic experience of widely used commercial software. They say that it would be "unprofessional" to release students onto the market without first-hand experience of major software. **Peter Coles**

Untimely death of Soviet virologist after hostile letters

London

DR VIKTOR Zhdanov, a pioneer of Soviet AIDS research, who died of a cerebral haemorrhage last July, may have had his end hastened by a campaign of hostile anonymous letters, according to *Sovetskaya Kultura*, a thrice-weekly newspaper published under the auspices of the Central Committee of the Communist Party of the Soviet Union. During the past three years, the paper says, a spate of such denunciations, addressed to various high-level Party and administrative bodies, accused Zhdanov of various academic malpractices, including the use of ghost-writers and "finding his wife a snug place" in his own institute.

Although these charges were eventually dismissed and although, in the words of one of Zhdanov's friends "it is impossible to associate his death with the anonymous letters in a juridical sense", the paper suggests that the harassment must have precipitated his collapse. It claims that its own investigations show that the attacks were written by disgruntled scientists who found it difficult to get on with Zhdanov.

Zhdanov was a forceful figure. Director of the Moscow Institute of Virology since 1961, he was one of the first Soviet scientists to take up the challenge of molecular biology and was the first to address the public on the problem of AIDS.

On the other hand, some of his own published results were, to say the least, controversial. A western colleague who visited him a few days before his death reported him to be in apparently excellent spirits, although he seemed to be concerned about his current project, an attempt to obtain money to establish a separate AIDS institute.

Sovetskaya Kultura seems, in part, to be using Zhdanov's death to highlight a loophole in Soviet law. Several years ago, the Soviet criminal code was amended to make anonymous libel an indictable offence, but there is as yet no provision in the law which would forbid officials who receive anonymous letters from considering and acting on them. Such an additional amendment, the newspaper considers should be introduced as soon as possible.

Vera Rich

Clinical research on the move

London

BRITAIN'S Medical Research Council (MRC) has declared itself in favour of abandoning its Clinical Research Centre for a physical and intellectual merger with the Royal Postgraduate Medical School. If and when that can be achieved, the council further intends to transfer its National Institute of Medical Research to form part of the new centre.

The council's decision, taken last week, is the latest shot in a battle that has raged since early last year, when the merger was first adopted as the solution to a "crisis in clinical research". Since then, the decision as to the site of the proposed centre has bounced between the council, a committee under Sir Robin Nicholson and a firm of management accountants. The council has now plumped for the Royal Postgraduate Medical School site, in the London district of Hammersmith, subject to some conditions that suggest the battle is not yet won.

As well as the obvious condition that sufficient land adjacent to the Hammersmith site can be acquired to make the merger possible, the council is not yet satisfied that the governance of the proposed centre, which would probably become a constituent school of the University of London, would allow the council sufficient financial and strategic control.

A further condition is that certain areas of research, notably psychiatry, infectious diseases and clinical molecular genetics, must be properly represented in the centre. These are seen as essential areas of clinical research for any national centre. Their extensive development at the Clinical Research Centre has been possible because of its physical association with a general community hospital. The Royal Postgraduate Medical School, by contrast, is not strong in these specialities and it would be both difficult and contrary to the current plans of the relevant health authorities to change that situation. "I do not see how anyone could decide on the Hammersmith site with expansion in mind", says Sir Christopher Booth, who will retire as director of the Clinical Research Centre by 1989, "because the health authorities have said there can be no expansion here". Yet another committee has been set to solve the problem.

The next step in the negotiations is for the Nicholson committee, on which all relevant parties are represented at a senior level, to make a definitive recommendation on siting. Even assuming it supports the Hammersmith site, it is estimated that it will take at least five years for the merger to be completed at a cost of more than £80 million. Peter Newmark

Plates rattle in Los Angeles earthquake

Washington

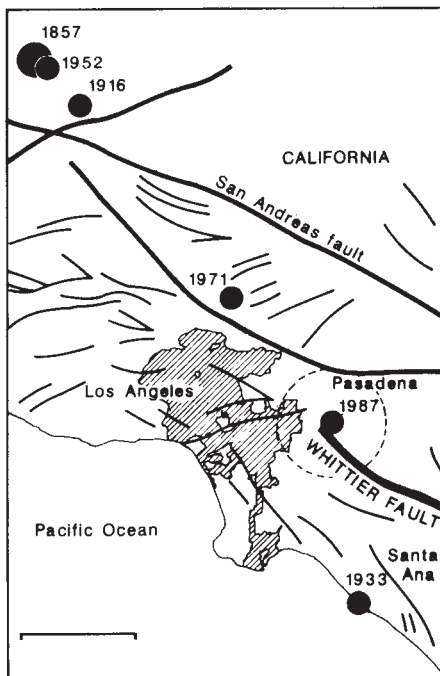
THE earthquake of magnitude 6.1 on the Richter scale which hit Los Angeles at 7.42 a.m. on 1 October, causing six deaths and widespread urban damage, and the aftershock on 4 October of magnitude 5.5, were unusual but not extraordinary events. Tremors of similar magnitude struck the metropolitan area in 1933 and 1971. What was unusual, according to Dr Lucy Jones of the California Institute of Technology, was that the preceding 11 months had seen no events larger than magnitude 4.0, the longest such interval recorded. Caltech's seismology instruments in Pasadena, nine miles north of the epicentre, were put out of action by the earthquake.

For the local population, the greatest significance of the quake was that it was on the Whittier fault, which is unconnected with the San Andreas fault about 30 miles away. Seismologists therefore see no portent in Thursday's event of any imminent catastrophe from the better-known fault system. For geologists, says Jones, the greatest scientific interest may derive from the earthquake's location at the end of the Whittier fault. Its dynamics could

shed light on how a fault line terminates.

Although many windows were broken and some freeway sections were cracked, damage was generally minor. The citizens of Los Angeles, well-drilled in earthquake procedure, gathered in the streets while the danger passed, and Mayor Tom Bradley was evidently pleased that the city coped efficiently.

David Lindley



The dotted circle shows the approximate area of damage of last week's earthquake. Large black circle, 8.0 on the Richter scale; smaller circles, 6.0-6.9 on the Richter scale. The black circle marked 1987 is the epicentre of the recent earthquake. Scale bar, 20 miles.

Lords on Europe's biotechnology

SIR—You have published (*Nature* 327, 650; 1987) an account of the UK House of Lords' report, *Biotechnology in the Community*, under the headline "European biotechnology plans leave the Lords cold". I find your report disappointing in that it neglects the quite positive support for many of the ideas expressed in the Commission discussion paper, COM(86)221, concerning agro-industrial development. Thus their lordships advocate "the extension of Community action in biotechnology and agro-industrial development into some of the suggested pilot projects".

This positive statement is accompanied by many helpful suggestions on the programme content, by a statement that the maximum funding of individual projects should be increased above the "very low level" available within earlier biotechnology programmes and by the advice (mentioned by your correspondent) that the Community should concentrate on projects not funded elsewhere and that are likely to show promise of economic viability.

The programme proposal we are now preparing will take into account their lordships' helpful comments, as well as the many other comments we have received and the 856 responses from European industry, research institutes and academic institutions that followed our call for expressions of interest in such a programme made last year.

The initial programme will be modest, in line with the reduced resources of the overall Framework Programme; but we are convinced that through launching support for a few worthwhile projects, we can promote increased land use and novel industrial development, utilizing the recent developments in biotechnology, and aiding the continuing evolution towards a market-driven agriculture.

P. FASELLA

Directorate-General for Science,
Research and Development,
Joint Research Centre,
Commission of the European Communities,
B-1049 Brussels, Belgium

Germ of an idea

SIR—We have twice^{1,2} been subjected to undocumented speculation by Hoyle and Wickramasinghe, who hypothesize that novel influenza virus variants come to us through seeding from comets. They state that the "epidemiological evidence is consistent with a model where the causative agent...is airborne and is brought down seasonally from a reservoir established in the stratosphere"². Why does the editor of

Nature not require Hoyle and Wickramasinghe to provide experimental evidence for their views? Where are the data showing that fragile lipid-containing influenza viruses survive extraterrestrial travel and when have influenza viruses been isolated from the stratosphere?

The data are equally consistent with simpler, more plausible explanations for which supporting data exist. For example, there is overwhelming sequencing evidence that influenza viruses undergo sequential evolutionary changes in nature with clear-cut ancestor-progeny relationships³. Furthermore, the appearance of novel influenza virus subtypes in man is clearly explained by a simple genetic exchange (reassortment of genes) of human and animal influenza viruses⁴, thus eliminating the need for seeding by comets. It appears untenable to ignore scientific evidence in favour of unsupported outer-space (spaced-out?) theories.

PETER PALESE

Department of Microbiology,
Mount Sinai School of Medicine,
1 Gustave L. Levy Place,
New York, New York 10029, USA

ROBERT WEBSTER

St Jude Children's Research Hospital,
332 N. Lauderdale, POB 318,
Memphis, Tennessee 38101, USA

1. Hoyle, F. & Wickramasinghe, N.C. *Nature* 322, 509–511 (1986).
2. Hoyle, F. & Wickramasinghe, N.C. *Nature* 327, 664 (1987).
3. Buonagurio, D.A. *et al.* *Science* 232, 980–982 (1986).
4. Webster, R.G. *et al.* *Nature* 296, 115–121 (1982).

But whose genome?

SIR—Word has reached those of us who work in the hinterlands of biology (on plants) that a project is afoot to sequence the human genome. Of course this raises ethical and practical questions. The one I like best is, "Whose genome is going to get sequenced?" May I make some suggestions? To begin with there is the problem of race, which is at least partly solved as it seems that at least one Asiatic and one Caucasian project are already in the works. The next question is that of nationality. It seems that every country that can afford it is going to have to have its own project. This line of reasoning continues until we get to the choice of the individual (R. Reagan, M. Thatcher, F. Mitterrand . . .) and then there are the social problems: class, accent and so on. As I see it, the human genome sequencing project will never get off the ground until a compromise can be found. My suggestion is that it go out to tender. This might help with the funding. Unfortunately, J.P. Getty and H. Hughes are dead, but there must be somebody who can afford to be sequenced.

DAVID TEPPER

Laboratoire de Biologie de la Rhizosphère,
Institut de la Recherche Agronomique,
78000 Versailles, France

Eiffel Tower in space

SIR—The protests against the still-evolving concept of an 'Eiffel Tower in space' (*Nature* 326, 125; 1987) are reminiscent of similar protests mounted against the original Eiffel Tower, the Statue of Liberty and, indeed, almost every monumental edifice that has risen above the horizontal. Recent complaints from the astronomical community may be more serious, but there are several mitigating factors.

(1) The proposed structure would be a circle ≤ 30 arc-minutes in diameter of star-like points of light resembling Vega in brightness and having an integrated visual magnitude ≥ -5 . It would be in the local sky at any given point on the Earth for only a few minutes of every night. Although the light from the sculpture might affect photometric astronomical observations in the rather improbable circumstance that it should visit the vicinity of the small regions of sky being observed, the general effect would be slightly greater than that of the planet Venus and much less than that of the Moon. As the orbit will be known, astronomers can easily plan their observations to avoid the problem.

(2) Although astronomical photographs showing 'contamination' by faint satellites are well known, these are usually wide-angle deep-survey pictures. There will be only one 'Eiffel Tower' structure, along with Soviet and US space stations and a few other man-made objects brighter than 1st magnitude.

(3) There is a lot of space, and the probability of collisions with useful satellites can be made ignorably small by judicious choice of orbit.

(4) The future of astronomy will be built in space, not on the Earth. Easy and inexpensive access to space, as provided by reusable aerospace planes and heavy-lift launch vehicles, will transform astronomy. The enormous scientific payoffs from the relatively modest space astronomy payloads launched so far amply demonstrate the importance of maintaining public interest in space development — interest that the 'Eiffel Tower' will certainly enhance.

Just as the original Eiffel Tower was a celebration of the progress and promise of engineering, so the new Eiffel Tower in space will be a celebration of the promise of space for the twenty-first century. I hope that this daring concept succeeds. I hope the critics will reflect on the beneficial interpretations and turn their attention instead to the real threats to the night sky such as air pollution and light pollution. The world can well use a new symbol of idealism and inspiration.

JOHN D. G. RATHER

Washington, DC, USA

Valence bonds for different people

Explanations of high-temperature superconductivity seem to be settling about the concept of the valence bond, which appears to mean different things to different people.

THE *rapprochement* between chemistry and physics stimulated by the search for an explanation of high-temperature (90 K) superconductivity evidently goes deeper than might have been expected. So much is plain from the material appearing in physics journals, but there is a long way to go before the *rapprochement* will be complete. For the time being, there seems to be a string of semantic issues to be resolved, people's understanding of what is meant by a valence bond, for example, a notion that seems all the more important now that P.W. Anderson seems to be winning people round to his view that resonating patterns of valence bonds are at the root of the new phenomena.

Part of the trouble is that chemists and physicists use different language to refer to the same concept and part of it is that that they also mean slightly different things when they say "valence bond". Some of the confusion stems from the way in which people are taught their crafts, but there is little doubt that what matters most to a chemist about a valence bond is that it should be strong enough to hold a molecule, or some part of it, together.

Physicists, however, talk as if all that matters about a valence bond is that it is a device for satisfying Pauli's exclusion principle — that two electrons cannot be in the same state at the same time unless their spins are antiparallel. This seems to be the reason why the words physicists use for the discussion of valence-bond patterns in solids are often simple adaptations of those used for the discussion of magnetism (antiferromagnetism as well as ferromagnetism). Each lattice point carries a spin, the energy of the interactions between a pair of neighbours is proportional to the product of the values of the two spins (with a constant of proportionality which is negative in the case of a ferromagnet). So does it not make sense to think of a valence-bonded solid structure as the ground state of a spin-loaded lattice in which the constant of proportionality is positive, favouring neighbouring antiparallel spins?

In many ways, the notation is simplifying. It has a great success in discussions of polyacetylene, for example, where if the backbone of the molecule is thought of as held together by inviolable sigma bonds between adjacent carbon atoms, the question is merely how to dispose of the spare pair of electrons on each carbon atom. But there are circumstances when the rep-

resentation of valence bonds as neighbour-neighbour spin couplings can be misleading. Although it has now emerged, under the impetus of the hunt for an explanation of superconductivity, that La_2CuO_4 is indeed an antiferromagnetic material, it is also the case that the nearest-neighbour interactions responsible for magnetic phenomena are not nearly as strong as those responsible for valence bonding.

More seriously, although the simple spin representation of one-dimensional polymers with spare electrons has neatly pointed to the importance of soliton solutions (say a break in a normally regular alternation of bond types), an unusual Chinese Japanese collaboration (Chang-qin Wo and Xin Sun from Fudan University, Shanghai, and Keiichiro Nasu from the Molecular Science Institute at Okazaki) has now shown how great are the complications of attempting to allow for electron-electron interactions even in simple one-dimensional polymers (*Phys. Rev. Lett.* **59**, 831; 1987).

It is remarkable that the group is able to say anything at all about the dependence of the alternation of bond lengths on the parameters defining the problem, the strength of the interaction between electrons and the lattice, or the range of their interaction with each other (which depends on the polarizability of the backbone bonds, among other things). Yet these problems may ultimately have to be faced by those who would use valence-bond models to account for high-temperature superconductivity.

With all that said, the formal spin-coupling representations of valence bond lattices do appear to be making some headway, most remarkably in an article by R.H. Parmenter of the University of Arizona (*Phys. Rev. Lett.* **59**, 923; 1987), who constructs a hamiltonian for a crystal lattice (or even two interpenetrating lattices) accommodating spin-spin interactions between nearest neighbours and between next-nearest neighbours by means of an intermediate atom (shades of the Cu-O-Cu chains in the yttrium copper oxide structures).

Parmenter concludes that, with a suitable choice of parameters, a lattice can be at once antiferromagnetic (demonstrated by long-range order) and superconducting, which contradicts one of P.W. Anderson's more recent conclusions (with G. Baskaran, Z. Zou and T. Hsu in *Phys. Rev. Lett.* **58**, 2790; 1987) that the electron

holes created by the movement of valence bonds are inimical to antiferromagnetism. But if there should be exactly one electron for each lattice site, the antiferromagnetism will cause the electron band to be filled and superconductivity will not be possible.

The argument by Anderson and his colleagues is interesting because it makes more explicit than in previous papers the presumed link between the phase transition in La_2CuO_4 from a tetragonal to an orthorhombic crystal structure (which takes place at 550 K for the undoped material, but at lower temperatures when barium is substituted for lanthanum) and the occurrence of superconductivity. The distortion of the tetragonal crystal structure, this view says, signals the arrival of a phase in which resonating valence bonds are dominant, with quite substantial fractions of an electron charge transferred to the intraplane oxygens (already supposed to be doubly negative). But in those circumstances, the square symmetry in the plane will be broken, making each a parallelepiped.

On this view, say Anderson and his colleagues, exactly the parameters that control the phase transition in crystal structure also control the transition to superconductivity at lower temperatures. In particular, they say, these parameters determine that the superconducting transition temperature should increase with the proportion of barium doping while the crystal transition temperature falls.

What will the chemists make of all this? They will probably acknowledge that the representation of interacting atoms by quantum spin operators is unavoidable, given that there is an infinite lattice to handle. The physical difference between the application of valence-bond methods to simple molecules such as that of hydrogen and to lattice structures is that, for the molecule, it is necessary to take account of bizarre configurations in which both electrons surround one or other of the nuclei, leaving the other bare. In a crystal, it is much easier to suppose that there should be a single extra charge, or a few of them, on some of the interacting atoms, which is another way of saying that there will be a plethora of modestly excited states that common sense suggests should be counted, which is what the formalism of spin operators does. The snag is that it will take some time before everybody tumbles to this.

John Maddox

MHC protein structure

Those images that yet fresh images beget

Alain Townsend and Andrew McMichael

It is becoming clear that the predominant function of the proteins encoded by the major histocompatibility complex (MHC) is to bind selected antigens and thus make them recognizable by T cells. But the nature of the interaction between antigen, MHC molecule and T-cell receptor has long been the subject of intense debate. The general reader will understand, then, why every immunologist's pulse will race as he or she sees the three-dimensional structure of the binding site of an MHC molecule displayed for the first time in the two papers by P.J. Bjorkman and colleagues on pages 506 and 512 of this issue^{1,2}.

It is 13 years since the importance of the MHC became apparent with the discovery that T lymphocytes can recognize antigen only in association with the MHC molecules on the surfaces of cells. With the solution of the crystal structure of the MHC molecule HLA-A2 reported by Bjorkman *et al.* in this issue, it will now be possible for the first time to understand the structural basis of this phenomenon.

Models

Soon after the discovery of MHC-restricted recognition³, two simple models were proposed to account for it. On the one hand, recognition of MHC and antigen might be independent events, mediated by two receptors, and on the other, a single T-cell receptor may recognize a complex of MHC with antigen. The second of these views has now been vindicated by the following two findings. First, for both class I- and class II-restricted T cells, a single receptor can be transferred from one T cell to another and confer dual specificity for MHC and antigen^{4,5}. Second, the peptide antigens recognized by class II-restricted T cells do bind selectively to isolated class II MHC molecules *in vitro* (with affinities in the micromolar range) to form complexes that stimulate T cells specifically^{6,7}. The latter experiment has yet to be done with class I molecules (like HLA-A2), but the result seems likely to run parallel as the epitopes recognized by class I-restricted T cells can be defined with short peptides just as for most T cells restricted through class II molecules.

One of the perplexing aspects of the binding of antigens by MHC molecules is the simple arithmetic of there being only a few polymorphic MHC molecules expressed in each individual, but a very large number of foreign and self antigens that are potential ligands. One way that this

paradox can be resolved is if each MHC molecule has many separate binding sites for antigens. But the finding that most peptides that bind the same class II molecule also show competition for binding implies the presence of a single site⁶⁻⁹. This interpretation is consistent with the first important feature revealed in the A2 structure reported in this issue — the existence of a single potential antigen-binding site. This site is located on the top surface of the molecule facing away from the plane of the membrane, and is defined by a deep groove running between two long α -helices derived from the $\alpha 1$ and $\alpha 2$ domains of the molecule. The floor of the groove is formed by β -strands also derived from both $\alpha 1$ and $\alpha 2$ domains.

The evidence most strongly suggesting that this groove in the structure elucidated by Bjorkman *et al.* is the binding site, is the finding that most amino-acid residues in class I molecules that are known to influence the recognition of antigen by T cells are clustered in the walls or floor of the groove. In addition, there is an intriguing unidentified occupant of the site that has co-crystallized with the A2 molecule.

The size of the groove is approximately 25 Å long, 10 Å wide and 11 Å deep, and would be expected to accommodate peptides of about 8–20 amino acids in length, depending on the extent to which they are compressed by coiling or bending. This estimate of the expected size of ligands fits well with the experimental evidence reported in ref. 10. A peptide of 12 amino acids derived from the influenza matrix protein is recognized efficiently by A2-restricted human cytotoxic T lymphocytes¹⁰. The five peptide epitopes so far described that are restricted through murine class I molecules have all been defined with peptides that are 14–16 amino acids long (listed in ref. 10). A systematic study of the maximum and minimum lengths of peptides retaining activity in class I-restricted T-cell recognition experiments has not yet been done. But in one of the above examples, reducing an active peptide from 14 to 11 amino acids in length reduces its activity by about 10⁴ fold. The optimum size of peptides recognized by class II-restricted T cells is also in the 9–20 amino-acid range. These results imply that most peptides occupying the binding site are to some extent compressed by folding.

If many different peptide ligands can occupy the binding site of a single MHC

molecule, one might expect that they would share some structural homology. If bound peptides assume a common secondary structure, this homology should be visible in their amino-acid sequences. This has been borne out to some extent by comparing sets of peptides known to bind to identified class II molecules⁶⁻⁹. There are cases, however, where some peptides within a set have no detectable similarity to the others¹¹. Although the data are limited, this result raises the possibility that several independent homologous sets might be able to bind in a given single binding site. The site in A2 seems large enough to entertain this idea, as peptides of, say, 12 amino acids in α -helical conformation would have room to associate in several different overlapping positions within the site. This could explain why a single binding site seems able to accommodate many ligands. It is hoped that it will now be possible to co-crystallize several unrelated ligands with the A2 molecule and analyse the nature of the binding in detail.

The ability to identify sequences of protein antigens likely to be restricted through known class I and class II molecules, and the possibility of designing high-affinity ligands would have important implications for the manipulation of the immune response to pathogens or autoantigens. It may be possible, for example, to use a high-affinity ligand to block recognition of selected MHC molecules by T-cell clones that are responsible for tissue damage. A speculative example might be inhibition of HLA-B27-restricted T cells in ankylosing spondylitis.

As their name implies, MHC proteins provoke strong T-lymphocyte-mediated responses resulting in graft rejection when they are not matched in tissue transplantation (a phenomenon called alloreactivity). It has been known for more than 20 years that in each individual there is a very high frequency of T lymphocytes that can respond to a single foreign MHC difference. About 1:10⁵ T lymphocytes in mice of the H-2b type, for example, respond to variant class I MHC molecules that differ in only three or less amino acids from the wild type¹². In comparison, less than 1:10⁵ T cells in unimmunized mice recognize influenza virus antigens¹³. Thus, a minute difference in the MHC protein makes it about 100 times more immunogenic than a viral antigen. The reason may lie in the nature of the unidentified peptide(s) bound in the crystal. If a set of

Corrigendum

IN discussion of the paper by R.A. Cross *et al.* (*FEBS Lett.* 219, 306; 1987), it was implied that, in regard to the experiments on association of caldesmon, it had not been specified whether or not reducing conditions had been used (*Nature* 328, 669; 1987). They were, and the difference from the results of Lynch *et al.* remains to be explained. □

peptide fragments from degraded self-proteins bind to a particular wild-type class I molecule, it is likely that a variant class I molecule will bind a different set (for an earlier version of this argument, see ref. 14). This is because the amino-acid changes in the variants occur in the binding site and are likely to change its specificity. Thus, a minimal change in sequence at a critical position in a variant MHC molecule could result in the formation of a large array of new complexes with endogenous peptides that would be recognized as foreign in an individual bearing the wild-type MHC.

One of the fundamental questions that is raised by the crystal structure of HLA-A2 reported in this issue is where and how in the cell do peptides bind? The observation¹⁵ that viral proteins newly synthesized in the cytoplasm can furnish ligands for class I molecules, but probably not for class II (discussed in two recent News and Views articles^{16,17}) implies that the peptide ligands destined for binding to class I and class II molecules originate in different compartments of the cell. Theoretically, binding could occur at any time between the folding of the MHC molecule (with formation of the binding site) and the appearance of the complete molecule at the cell surface. An attractive hypothesis is that peptides bind class I molecules as they fold and associate with β_2 microglobulin. In this case, peptides might survive in place for the lifespan of the MHC molecule. This suggestion is compatible with the slow rate of dissociation of peptide from class II molecules measured *in vitro*⁷, and for the surprisingly long time (up to 72 h) that a cell remains a target for specific CTL after a short exposure to a high concentration of peptide antigen (our unpublished results). This problem is one of many that are highlighted by the solution^{1,2} of the crystal structure of HLA-A2. T-cell recognition can now be explored in a rational way. □

1. Bjorkman, P.J. *et al.* *Nature* **329**, 506–512 (1987).
2. Bjorkman, P.J. *et al.* *Nature* **329**, 512–518 (1987).
3. Zinkernagel, R.M. & Doherty, P.C. *Adv. Immun.* **27**, 51–177 (1979).
4. Dembic, Z. *et al.* *Nature* **320**, 232–238 (1986).
5. Saito, T., Weiss, A., Miller, J., Norcross, M.A. & Germain, R.N. *Nature* **325**, 125–130 (1987).
6. Babbitt, B.P., Allen, P.M., Matsueda, G., Haber, E. & Unanue, E.R. *Nature* **317**, 359–360 (1985).
7. Buus, S. *et al.* *Cell* **47**, 1071–1077 (1986).
8. Guillet, J.G., Lai, M.Z., Binner, T.J., Smith, J.A. & Geffert, M.L. *Nature* **324**, 260–262 (1986).
9. Sette, A. *et al.* *Nature* **328**, 395–399 (1987).
10. Gotch, F., Rothbard, J., Howland, K., Townsend, A. & McMichael, A. *Nature* **326**, 881–882 (1987).
11. Allen, P. *et al.* *Nature* **327**, 713–715 (1987).
12. Widmer, M.B. & MacDonald, H.R. *J. Immun.* **124**, 48–51 (1980).
13. Askonas, B.A. *et al.* *Immunology* **45**, 79–84 (1982).
14. Matzinger, P. *Nature* **292**, 497–501 (1981).
15. Morrison, L.A. *et al.* *J. exp. Med.* **163**, 903 (1986).
16. Germain, R.N. *Nature* **322**, 687–689 (1986).
17. Bevan, M.J. *Nature* **325**, 192–194 (1987).

Alain Townsend and Andrew McMichael are in the Nuffield Department of Clinical Medicine, John Radcliffe Hospital, Headington, Oxford OX3 9DU, UK.

Superconducting ceramics

Finding characteristic lengths

Ted Forgan

ACCORDING to standard theories, the response of superconductors to magnetic fields is determined by two characteristic microscopic lengths: the penetration depth λ of magnetic field into a superconductor; and the coherence length ξ , the width of the narrowest possible interface between a superconducting and a normal region. These are of great practical importance: ξ controls the value of the highest field at which the material remains superconducting; and both influence the 'pinning' of magnetic flux lines which determines the highest current a superconductor can carry in a magnetic field without dissipation. R. Felici *et al.* on page 523 of this issue¹ report measurements of λ for the new high-temperature superconductors. These differ from previous measurements, however, and it is turning out to be remarkably difficult for investigators to agree on values for either λ or ξ in the new materials.

To understand the effect of magnetic fields on superconductors, one should bear in mind that the energy changes causing the superconducting state are extremely small (even in the new materials). A superconductor in a weak magnetic field excludes the magnetic flux (the Meissner effect) but this increases its energy — it is like a bar magnet aligned the wrong way in a magnetic field. For type-I superconductors (Fig. 1a), this increase in energy results in the disappearance of superconductivity above a critical field H_c , which is typically much less than 1 tesla (T). For type-II superconductors (Fig. 1b, c) the Meissner effect occurs only up to the lower critical field H_{c1} . Above this field, magnetic flux enters the material in the form of flux lines, each of which carries one quantum of flux, $\Phi_0 = h/2e \sim 2 \times 10^{-15}$ weber (ref. 2). The entry of flux reduces the magnetic energy and

allows the materials to remain superconducting until an upper critical field H_{c2} , which may be much larger than H_c , is applied. High-temperature superconductors appear to be type-II materials with H_{c2} much greater than 100 T.

This type-I/type-II behaviour can be explained within the framework of the standard Ginzburg–Landau theory³, which shows the importance of the ratio $\kappa = \lambda/\xi$. At small fields, when a superconductor is in the Meissner state, an applied magnetic field is not totally excluded but penetrates a very short distance λ below the surface. If at larger fields the flux lines are formed, each one has a 'normal core' of radius $\sim \xi$, surrounded by a vortex of supercurrents, that allows the magnetic field to extend to a radius $\sim \lambda$. If λ is larger than $\xi/\sqrt{2}$, the lowering of the magnetic energy can exceed the raising of the energy caused by the loss of superconductivity in the normal core, and type-II behaviour results. Flux lines enter at H_{c1} and form a lattice, with the lattice parameter related to the flux density within the specimen.

Near H_{c1} , the flux lines are roughly λ apart; as the field is increased, they move closer together until at H_{c2} they are roughly ξ apart. At this field, the normal cores overlap and superconductivity disappears. The formulae arising from the theory are given in Fig. 1 legend. In the simplest versions of this theory, λ and ξ are independent of field, and κ is independent of temperature. The values of these lengths are constant at low temperatures but they increase together to infinity at the critical temperature T_c , where the critical fields are all zero.

It might seem a simple matter to determine these microscopic lengths λ and ξ , by measuring the macroscopic magnetic properties of type-II superconductors

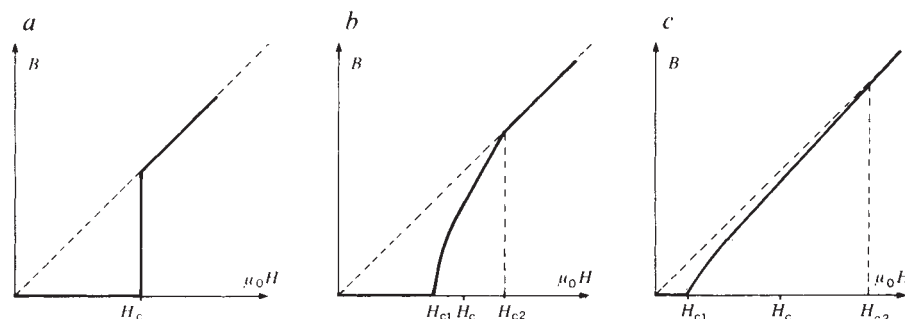


Fig. 1 Average flux density B inside a superconductor specimen as a function of applied field $\mu_0 H$ (ideal behaviour for a long rod parallel to field). *a*, Type-I superconductor, $\kappa < 1/\sqrt{2}$; for $H \geq H_c$ the sample becomes normal, with $B = \mu_0 H$. *b*, Type-II superconductor, $\kappa \geq 1/\sqrt{2}$; magnetic flux begins to penetrate at $H = H_{c1}$, and the sample becomes normal at H_{c2} . *c*, Extreme type-II superconductor, κ large. The equations from Ginzburg–Landau theory are: for type II, $\mu_0 H_{c2} = \sqrt{2} \kappa \mu_0 H_c = \Phi_0 / 2\pi \xi^2$; extreme type II, $\mu_0 H_{c1} = \mu_0 H_c \ln(\kappa) / \sqrt{2} \kappa = \Phi_0 \ln(\lambda/\xi) / 4\pi \lambda^2$.

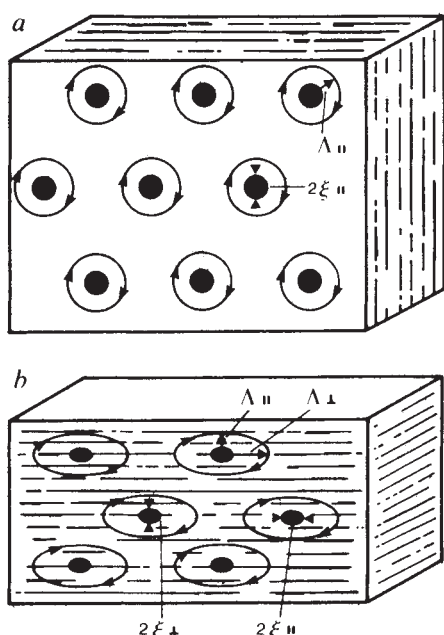


Fig. 2 Flux-line lattices in an anisotropic superconductor; the applied field is directed into the page. *a*, Applied field perpendicular to the Cu–O planes. H_{c1} and H_{c2} depend only on $\xi_{\parallel}$ (radius of the flux core) and $\lambda_{\parallel}$ (radius of the supercurrent vortex). *b*, Applied field parallel to Cu–O planes. H_{c1} and H_{c2} depend on $\xi_{\parallel}$, $\xi_{\perp}$, $\lambda_{\parallel}$ and $\lambda_{\perp}$. Note $\xi_{\parallel}/\xi_{\perp} = \lambda_{\perp}/\lambda_{\parallel}$.

and applying the formulae. However, in high- T_c materials there is a complication arising from their layered crystal structure⁴. It is much easier for supercurrents (and normal currents) to flow in the Cu–O planes than perpendicular to them. Hence λ is anisotropic, being shorter for fields controlled by supercurrents flowing in the planes. The coherence length ξ is also anisotropic, so that the shape of the flux-line lattice, the normal cores, the values of κ and the critical fields all depend on the field direction. The anisotropic Ginzburg–Landau theory for this case was worked out previously for layered conventional superconductors⁵ (see Fig. 2). In a polycrystalline sintered sample, which is the form most high- T_c samples have taken up till now, any measurements of these quantities will be some weighted average of their values for various crystal orientations.

What are the methods of measuring λ and ξ in high- T_c superconductors at low temperatures? Some data on single crystals, obtained using magnetic measurements, have recently been published⁶ and the results look nothing like the ideal case shown in Fig. 1. This is undoubtedly because of the pinning of flux lines at surfaces or defects, so that they do not form an equilibrium distribution. This is good for current-carrying applications, but not for investigation of fundamental properties. As a result, H_{c1} does not appear as a clean break on a magnetization curve, but is taken as the first point on the curve where deviation from

Meissner-type behaviour is seen.

The values of H_{c2} at low temperature are well above the steady fields available in most laboratories, so they are inferred by extrapolation from measurements near T_c . Also, the disappearance of superconductivity at H_{c2} has to be detected. This may be affected by superconductivity in the surface of the samples (which can exist in a thin layer up to $1.69 H_{c2}$; ref. 3).

Worthington *et al.*⁷ use the values of H_{c2} in two crystal directions and of H_{c1} in one of them to deduce the lengths given in the table. The other value of H_{c1} that they measure is inconsistent with these values — if it were used instead, the values of ξ would be the same but the values of λ would be increased by a factor of about 7. It is noteworthy how close the coherence length perpendicular to the Cu–O plane, $\xi_{\perp}$, is to atomic dimensions. This means that the theory is being extrapolated very far from its usual realm of validity.

Muon-spin rotation (μ SR), a method that relies on the stopping of spin-polarized muons in a sample in a magnetic field, can also be used to measure λ . Muons decay in about 2 μ s, emitting a positron preferentially along one direction of their spin axis. If the muon spins are perpendicular to the magnetic field, they precess and the probable direction of positron emission sweeps round like a lighthouse beam at a rate determined by the magnitude of the magnetic field. A positron detector at a fixed angle gives a μ SR spectrum like that shown in Fig. 3. If the field varies with position inside the sample, then different muons precess at different rates, get out of step, and the oscillations become damped. Thus the muon is a microscopic field probe: the frequency of the μ SR signal gives the magnitude of the field, the damping gives the spatial variation and the amplitude shows whether the direction of the field is constant throughout the sample.

To measure λ in a superconductor by this technique, a field above H_{c1} should be applied, so that there is an inhomogeneous field from the flux-line lattice

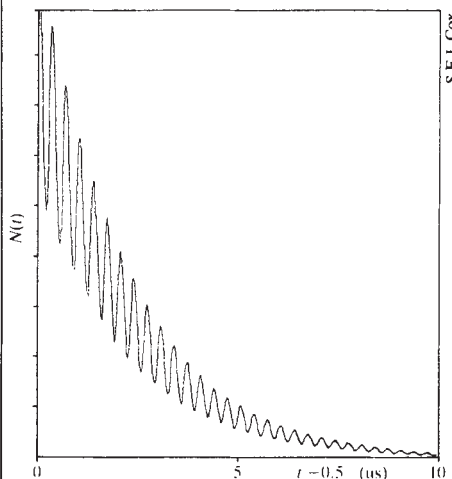


Fig. 3 Typical muon-spin resonance spectrum.

Length scales (Å) in YBa ₂ Cu ₃ O _{7-δ} at low temperatures*				
Method	$\xi_{\perp}$	$\xi_{\parallel}$	$\lambda_{\perp}$	$\lambda_{\parallel}$
Magnetic properties	7	34	260	1,250
μ SR [†]				4,500
μ SR [‡]				1,400
Polarized neutron reflection ¹				225
Small-angle neutron scattering ¹²				21

* All important lengths are perpendicular to the applied field direction. $\xi_{\perp}$ and $\xi_{\parallel}$ are the coherence length perpendicular and parallel to Cu–O planes. $\lambda_{\parallel}$ and $\lambda_{\perp}$ are the penetration depth with the supercurrent flowing parallel and perpendicular to the planes (see Fig. 2.)

† P. Birrer *et al.* (personal communication). ‡ D. Harshman *et al.* (personal communication).

within the specimen. It has been shown that for isotropic, large- κ superconductors (not too close to H_{c1} or H_{c2}) the damping rate of the μ SR signal should be proportional to Φ_0/λ^2 (ref. 9). This method has been used by at least two groups to obtain values for λ in polycrystalline samples of YBa₂Cu₃O_{7-δ} (see table). Before comparing them with the earlier values, we should consider the effect of anisotropy. For a field applied parallel to the planes, the effective λ measured by μ SR is the geometric mean of the parallel and perpendicular values, $\lambda_{\parallel}$ and $\lambda_{\perp}$. For the field perpendicular to the planes, the shorter value $\lambda_{\parallel}$ should be obtained. With a polycrystalline sample, the average value of λ should be a little less than the geometric mean. There is clearly a discrepancy between the two μ SR results, and between them and the magnetization results. There may be minor doubts about the μ SR technique: for instance, whether the muons are trapped randomly or at defects in the material, but I suspect that the main difference between different reports is in the samples rather than the technique.

Polarized-neutron reflection, the technique reported by R. Felici *et al.* in this issue¹, relies on the fact that when neutrons of wavelength λ are partially reflected at a grazing angle θ from the surface of a solid, the reflection coefficient is determined by any changes in neutron energy that occur within a distance $\sim \lambda/\sin \theta$ from the surface (~ 800 Å in the present case). In addition to the average energy of interaction between the neutrons and the nuclei in the solid, in the presence of a magnetic field there is also the Zeeman energy of the neutron, the spin of which is aligned either with or against the field. For a superconductor in the Meissner state, the Zeeman energy dies away over a distance λ . If λ is much larger than 800 Å (in the present case),

the reflection coefficient should depend little on the neutron spin direction.

This method sounds most attractive, as it relies on a basic definition of A , and has worked well in conventional superconductors¹². Once again, however, we have to worry about anisotropy. Without access to detailed computations, I would guess that the technique gives an average weighted towards the shorter values. If so, the value recorded in the table is a little short compared with the magnetization results and is definitely inconsistent with the μ SR values. In this case, the very directness of the reflection method may be a disadvantage as it measures a surface property of a sample, whereas μ SR looks at the bulk — but which is intrinsic? Also if there happened to be a nonmagnetic overlayer of just the right thickness to act like an anti-reflection coating, it could enhance the spin-dependence of the reflectivity¹³, which would look like a shorter A . The other neutron-scattering value for A in the table is obtained by a small-angle scattering technique. The interpretation of the data may be affected by inelastic scattering. The value is almost certainly incorrect.

So what can we conclude? Certainly that more work is needed — there are still some techniques, such as a.c.-field penetration and Bragg diffraction of neutrons by the flux-line lattice, that have not yet been reported. At least one of the techniques I have reviewed is probably giving misleading answers; the Ginzburg–Landau picture, especially its anisotropic form, might not be applicable to high- T_c superconductors at low temperatures.

It is a truism among workers in the field that experiments on single crystals are the answer. But, when made at high temperatures under atmospheric pressure, single crystals of the yttrium–barium material are oxygen deficient, tetragonal and not superconducting. To make them superconducting, they must be changed to orthorhombic by diffusing in oxygen. This may take 10,000 times longer for a 1-mm crystal than a 10- μ m powder. Whatever technique is used, can I make a plea for those performing measurements in the vortex state to apply the field above T_c and cool down in constant field? It is well established that in samples with pinning, this gives a better quality flux-line lattice with a flux density close to that applied. $\square$

1. Felici, R., Penfold, J., Ward, R.C., Ols, E. & Maticcotta, C. *Nature* **329**, 523–525 (1987).
2. Gough, C.E. *et al.* *Nature* **326**, 855 (1987).
3. Tilley, D.R. & Tilley, J. *Superfluidity and Superconductivity* (Hilger, Bristol, 1986).
4. David, W.I.F. *et al.* *Nature* **327**, 310–312 (1987).
5. Morris, R.C. *et al.* *Phys. Rev. B* **5**, 895–901 (1972).
6. Dinger, T.R. *et al.* *Phys. Rev. Lett.* **58**, 2687–90 (1987).
7. Worthington, T.K. *et al.* *Phys. Rev. Lett.* **59**, 1160–3 (1987).
8. Aeppli, G. *et al.* *Phys. Rev. B* **35**, 7129–32 (1987).
9. Lynn, J. *et al.* *Phys. Rev. B* (in the press).
10. Fletcher, G.P. *et al.* *Phys. Rev. Lett.* **52**, 1539–41 (1984).
11. Bland, J.C. *et al.* *Phys. Rev. Lett.* **58**, 1244–47 (1987).

Ted Forgan is in the Physics Department, Birmingham University, PO Box 363, Birmingham B15 2TT, UK.

Birds learn song from aggressive tutors

It is a well-established fact that young male song birds (members of the sub-order oscinae) acquire the typical song of their species by copying the songs of older individuals. This capacity for imitative song learning has been exploited by generations of bird fanciers, both in Europe¹ and the Orient², who have taught birds of various species to whistle elaborate songs or even folk tunes. The scientific study of song learning did not begin until the technology for sound analysis became available in the 1950s, and only now, in studies such as those by Clayton³, are the social factors that influence song learning being elucidated.

In their classic studies, W.H. Thorpe⁴ and P.R. Marler⁵ raised young males in acoustic isolation and taught them with tape-recorded song under a controlled regime. These studies gave rise to two key concepts in song learning: the sensitive

Clayton³ has recently investigated. She comes to the interesting conclusion that young males learn from the adults who are most aggressive towards them.

Clayton performed two experiments on zebra finches (*Taeniopygia guttata*; see figure). These birds are normally raised by both parents until they are about 35 days old, when they reach independence. In one of Clayton's experiments, young zebra finches were raised only by their mother, and at independence were transferred to a cage with two live male tutors. Social interactions between the young and the tutors were monitored and when the young birds were 4 months old, their own songs and those of their tutors were recorded. Each of the 11 birds in this experiment were found to learn their song from the tutor that was most aggressive (measured by the number of pecks and chases) towards it.

In another experiment, Clayton analysed the role of the paternal song in the young male's choice of which tutor to copy. Chicks were raised by both parents until independence and then placed in a cage with two tutors. In this experiment the three birds were separated by wire mesh, so that physical attack, shown to be important by the experiment described above, could not influence which bird the pupil copied. The tutors of

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Photo: Hans Reinhard (Bruce Colman).

period and the inborn template. The first refers to the finding that young birds can learn from tape recordings only during a certain period in life — between 10 and 50 days in white crowned sparrows (*Zonotrichia leucophrys*), for example; whereas the second refers to the fact that pupils are not equally likely to copy from any taped song: they seem to have an inbuilt predisposition to learn their own species' song or at least sounds that are very similar to it.

In the past few years, however, it has become clear that the results of song tutoring can be very much influenced by the experimental method used. A crucial modification in recent experiments is the use of live teachers instead of tape-recorders. When young birds can see and interact with a tutor, they are not only able to learn new songs well beyond the traditional sensitive period, but they are also willing to learn the songs of other species⁶. These results demonstrate that social interaction is important in determining the duration and nature of song learning, but they do not show what features of social interaction are important, and this is the question that

each young male were chosen so that one had a song similar to the pupil's father and the other did not. Nine of the ten birds learnt their songs from the tutor with a song like the pupil's father, and the tenth bird learnt from both tutors. Thus, the early experience of hearing their father singing results in a young bird's subsequent tendency to learn from a tutor that sounds similar.

These experiments are among the first to investigate how specific kinds of social experience influence song learning⁷. They help to elucidate both the complex process of behavioural development and the fascinating problem of how and why local dialects of bird song arise.

Dan Weary and John Krebs

1. *The Bird Fancier's Delight* (Meares, London, 1917).
2. Konishi, M. A. *Rev. Neurosci.* **8**, 125–170 (1985).
3. Clayton, N. *Anim. Behav.* **35**, 714–721 (1987).
4. Thorpe, W.H. *Ibis* **100**, 535–570 (1958).
5. Marler, P.R. *J. comp. physiol. Psychol. Monogr.* **72**, 1–25 (1970).
6. Baptista, L.F. & Petrinovich, L. *Anim. Behav.* **32**, 172–181 (1984).
7. Payne, R.B. *Anim. Behav.* **29**, 688–697 (1981).

Dan Weary and John Krebs are in the Edward Grey Institute of Field Ornithology, Department of Zoology, South Parks Road, Oxford OX1 3PS, UK.

Star formation

To B or not to B ?

Alan P. Boss

STAR formation is an astronomical puzzle that involves deciphering how interstellar molecular clouds turn into stars. The puzzle lies in the fact that both gravitationally bound, dense molecular clouds and newly-formed pre-main-sequence stars can be observed, but precious little can be seen in between. What are the intermediate objects, and what physical processes dominate their evolution? Two recent papers^{1,2} address the second question: specifically, the importance of magnetic fields for star formation. Crutcher, Kazès and Troland¹ use radio observations of Zeeman splitting in OH molecules to argue that magnetic fields could be dynamically important for dense clouds and star formation. Working at the other end of the puzzle, Tayler² has re-interpreted measurements of chromospheric activity in pre-main-sequence stars as evidence of 'fossil' magnetic fields, and estimates the strength of the magnetic field necessary at the beginning of convective evolution to explain the present chromospheric activity.

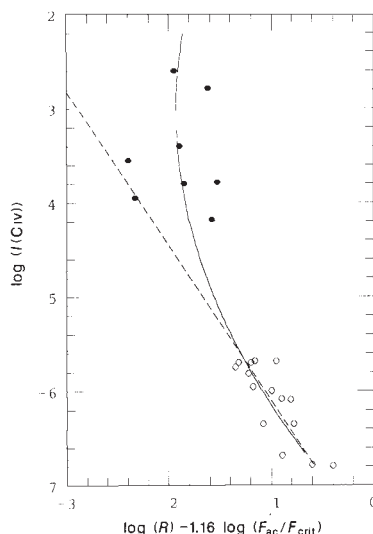
Knowledge of the magnetic field strength (B) in dense molecular clouds is essential for determining to what extent magnetic fields are dynamically significant during protostellar contraction and collapse. Most calculations of protostellar collapse have assumed $B = 0$, for reasons of simplicity as much as for the belief that at some point magnetic fields dynamically decouple from the gas, through processes such as ambipolar diffusion of ions through the neutral bulk of the gas. The acid test is simply to measure B in star-forming regions.

Measurements of B in dense molecular clouds are best accomplished by looking for the Zeeman effect, in which an external magnetic field splits an atomic state with total angular momentum J into $(2J + 1)$ energy levels. For molecular clouds, where B is of the order of microgauss (μG), the frequency splitting is very small, a few hertz, whereas molecular lines are broadened by Doppler effects by kilohertz. The Zeeman effect must be estimated from the difference in intensity between the right- and left-handed circularly polarized portions of the radiation; the difference is directly proportional to the magnitude of the component of the field parallel to the line-of-sight ($B_{\parallel}$).

Zeeman splitting has been detected in both the 21-cm spectral line of neutral hydrogen (H I) and the 18-cm OH lines. Because the OH lines sample considerably denser regions than the H I line, Crutcher, Kazès and Troland have looked

for OH Zeeman splitting in four star-forming molecular clouds. Their estimates of $B_{\parallel}$ range from 14 μG for the W40 cloud to 69 μG for the S88B cloud. Both these clouds have already undergone the formation of massive (OB) stars. Added to the five previous detections of OH Zeeman splitting by these authors and by Heiles and colleagues^{3,4}, the results suggest¹ that on average, B is about 30 μG in cool clouds not heated by OB-star formation, whereas it is about 120 μG in hot clouds containing OB stars. This difference can be attributed to the higher densities being sampled in the warmer clouds: theoretically one may expect the field strength to increase with gas density⁵. Roughly a dozen other clouds have been similarly observed, but have yielded only upper limits of 15–30 μG (ref. 6).

How important is a field of this strength for the dynamics of a dense cloud that has not yet formed stars? Unfortunately, this is very hard to answer without knowing the gas density corresponding to the measured field strengths. One way of trying to answer this question is to examine the ratio δ of the magnetic-field energy to the self-gravitational energy: $\delta \sim 1$ implies that magnetic fields dominate the cloud dynamics. Myers and Goodman (personal communication) have evaluated a sample



Stellar chromospheric activity (measured by the strength of carbon emission — $I(\text{C iv})$ — normalized to the total stellar luminosity) as a function of the Rossby number R and the acoustic flux F_{ac} (normalized by $F_{\text{crit}} = 10^7 \text{ erg g}^{-1} \text{ s}^{-1}$). Pre-main-sequence stars (open circles) are considerably more active than main-sequence stars (filled circles). Solid line, Gilliland's¹⁰ quadratic fit; dashed line, Tayler's² proposal for pre-main-sequence stars with negligible fossil magnetic fields. (From refs 2, 10.)

of more than 100 molecular clouds, under the assumption that the observed line widths are caused by magnetic motions, including Alfvén waves, and thermal velocity dispersion. They find that a field of B of about 30 μG (together with thermal pressure at a temperature of 10 K) is sufficient to support *all* of the clouds against collapse. Assuming equilibrium, for clouds larger than about 10 parsec (pc), δ is about 1, but drops as low as about 0.1 for smaller clouds. Evidently, low-mass clouds have experienced significant loss of magnetic flux. Myers and Goodman argue that ambipolar diffusion is responsible.

Another method, although indirect, of determining the dynamical importance of magnetic fields is to compare the optical polarization produced by dust grains aligned by the magnetic field with observations of the axes around which clouds are flattened and/or rotating. Heyer⁷ has done this for the dense cores in the Taurus molecular cloud, and finds that the dynamically chosen axes are uncorrelated with the direction of the magnetic field, implying that magnetic fields do not dominate the dynamics of these cores. Thus, it seems that molecular clouds are primarily supported by magnetic fields on the scale of 10 pc and larger, but magnetic fields are less dynamically significant on the smaller scale (0.1–1 pc) of the most dense cloud cores.

One can also hope to pin down the importance of magnetic fields for star formation through observations of B in pre-main-sequence stars (T Tauri variable stars). The Zeeman effect has been detected in optical lines from certain peculiar main-sequence stars, but so far observations of T Tauri stars have yielded only upper limits of the order of hundreds of gauss^{8,9}. The alternative is to infer the presence of magnetic fields through the intensity of emission from stellar chromospheres: magnetic and acoustic waves are thought to be necessary to feed energy to the chromosphere from the underlying photosphere.

Young stars show considerably greater chromospheric activity than main-sequence stars (see figure). Gilliland¹⁰ has shown that chromospheric activity is a function of the ratio of the stellar rotational period to the timescale for convective overturn (the Rossby number) and the amount of acoustic energy deposited in the chromosphere; the former dependence was expected on the basis of mean-field dynamo theory¹¹. Although Gilliland has tried to fit pre-main-sequence stars into the same picture as main-sequence stars, Tayler² offers an intriguing alternative explanation. He hypothesizes that low-mass protostars arrive at the beginning of their convective evolution tracks with substantial fossil magnetic fields (between 10^{-3} and 10 G), remnants of the

magnetic flux contained in the original molecular cloud. Fields of this strength would dominate dynamo-produced fields through much of pre-main-sequence evolution, until the fossil fields decay through ohmic or turbulent resistivity. The chromospheric activity would then fall to values expected from dynamo theory. Tayler's hypothesis can be checked by measuring B in young stars; although no detections have yet been made, the measured upper limits on B are still far above Tayler's hypothetical range.

When firm values for B in young stars are determined, the amount of magnetic flux lost during collapse from molecular clouds can be estimated. If frozen-in magnetic fields increase with density⁵ as $n^{1/3}$ – $n^{1/2}$ throughout protostellar collapse, a 30 μ G field in a dense cloud core could be

amplified to between 30 and 30,000 G in a young star. The latter value has already been excluded for three T Tauri stars^{8,9}, but the challenge of detection remains. $\square$

1. Crutcher, R.M., Kazès, I. & Troland T.H. *Astr. Astrophys.* **181**, 119–126 (1987).
2. Tayler, R.J. *Mon. Not. R. astr. Soc.* **227**, 553–561 (1987).
3. Heiles, C. & Stevens, M. *Astrophys. J.* **301**, 331–338 (1986).
4. Troland, T.H. & Heiles, C. *Astrophys. J.* **301**, 339–345 (1986).
5. Mouschovias, T.C. *Astrophys. J.* **207**, 141–158 (1976).
6. Crutcher, R.M., Troland, T.H. & Heiles, C. *Astrophys. J.* **229**, 134–137 (1981).
7. Heyer, M.H. *Astrophys. J.* (in the press).
8. Johnstone, R.M. & Penston, M.V. *Mon. Not. R. astr. Soc.* **219**, 927–941 (1986).
9. Johnstone, R.M. & Penston, M.V. *Mon. Not. R. astr. Soc.* **227**, 797–800 (1987).
10. Gilliland, R.L. *Astrophys. J.* **300**, 339–347 (1986).
11. Durney, B.R. & Latour, J. *Geophys. Astrophys. Fluid Dyn.* **9**, 241–255 (1978).

Alan P. Boss is in the Department of Terrestrial Magnetism, Carnegie Institution of Washington, 5241 Broad Branch Road NW, Washington DC 20015, USA.

Reproductive biology

Causes of death before birth

Jared M. Diamond

OF human pregnancies recognized by a missed menstrual period, about 15% end in miscarriage. When introduction of hormone assays permitted earlier detection of pregnancies that would otherwise have gone unrecognized, an even higher fraction of those pregnancies was found to terminate without issue, yielding a combined miscarriage rate of 50% (ref. 1). Efforts at artificial fertilization of ova within fallopian tubes point to further losses of conceptuses before implantation, suggesting overall miscarriage rates as high as 80% (ref. 2). What are the causes of this prenatal carnage? Recent studies by Vadheim and colleagues^{3,4} are triply interesting in identifying an unappreciated cause of miscarriage, an explanation for the persistence of a lethal genetic disease, and a role for natural selection at the prenatal stage.

Just as male psychiatrists used to blame most emotional problems of children on defective postnatal mothering, male biologists used to attribute most miscarriages to prenatal maternal factors, such as uterine physical abnormalities, infections, endocrine dysfunction and physical trauma. These factors are now thought to be responsible for only about 15% of miscarriages. Chromosomal structural abnormalities cause about 50%, especially during the first half of gestation^{1,5,6}. In particular, cases of trisomy (triploid chromosomes resulting from non-disjunction) in spontaneous abortuses have now been reported for each human autosome except numbers 1 and 17, although only trisomies 21 (Down's syndrome), 13 and 18 are likely to survive birth. (An extra chromosome number 1 has been seen in sperm⁷ but is

evidently too lethal for a conceptus to survive to an age at which an abortus can be recovered.) Although some remaining miscarriages involve blood-group incompatibility, a large unexplained residue still remains.

Vadheim *et al.* have looked at insulin-dependent diabetes mellitus (IDDM, formerly called juvenile-onset diabetes), a genetic disease that was lethal until the advent of insulin-replacement therapy. The primary diabetogenic genes probably reside in the region of the HLA gene complex (major human histocompatibility complex), and the strongest association of IDDM is with the HLA alleles *DR3* and *DR4*. An individual inheriting *DR3* from one parent and *DR4* from the other bears the highest risk of diabetes.

From a pool of families in which at least one child was known to have IDDM, Vadheim *et al.* examined all children of those couples in which either the mother or the father was heterozygous for the diabetogenic alleles *DR3* and *DR4*. In general, one expects a parent heterozygous for a certain allele to transmit that

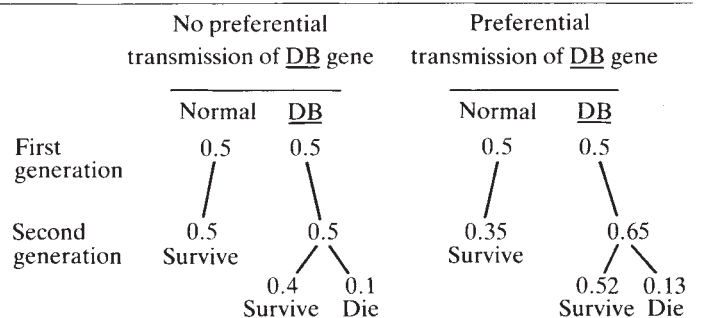
allele to 50% of his/her offspring and to transmit his/her other allele to the other 50% of offspring. In fact, *DR3* is preferentially transmitted by both the fathers and the mothers, and *DR4* by the fathers but not the mothers. That is, 68, 65 and 72% of the children inherit a paternal *DR3*, maternal *DR3* and paternal *DR4* allele (significantly different from the expected 50% at the $P < 0.002$ level or better), but only 56% inherit the maternal *DR4* allele (not significantly different from the expected 50%). Although apparently preferential transmission of diabetogenic alleles is expected in families chosen for having a diabetic child, this bias fails to explain the observed preferential transmission to the non-diabetic children as well. That is, even non-diabetic children born alive are more likely to receive the diabetogenic allele than a normal allele.

Could the explanation be gamete selection — that sperm with *DR3* or *DR4*, or ova with *DR3*, are more likely to result in conceptuses than gametes with the normal alleles? Or could the explanation instead be intra-uterine selection — conceptuses with *DR3* or *DR4* are less likely to miscarry than conceptuses with the normal alleles? Several types of evidence favour the latter interpretation. For example, when the families were divided into those with and without a history of spontaneous miscarriages, the former group far exceeded the latter in the proportion of couples composed of a *DR4* father and *DR3* mother (71 rather than 28%), the combination most likely to produce a diabetic child. The preferential transmission increased with the birth of each diabetic child, implying that mothers who had already carried a diabetic child were increasingly likely in the next pregnancy to miscarry fetuses having normal alleles compared with those having diabetogenic alleles.

Various relationships between parental haplotype and risk of miscarriage are well known for other HLA alleles^{8,9}. Neither in these other cases nor in the case of *DR3* and *DR4* discussed above is the mechanism of miscarriage known, but it presumably involves an error in the antigenic signalling that controls whether a pregnant mother's immune system will reject the foreign body that is the fetus.

If diabetogenic (*DB*) and normal alleles have equal transmission probabilities, the former allele will decrease in frequency from generation to generation and will be eliminated by selection, as 20% of its carriers die if untreated (left). But if there

is 65:35 preferential transmission, the diabetogenic allele frequency will be maintained despite the 20% postnatal mortality (right). (Based on ref. 3.)



Whatever the mechanism turns out to be, the observations of Vadheim *et al.* are relevant to several problems.

First, the incidence of insulin-dependent diabetes is much higher in offspring of diabetic fathers than of mothers (5–6% against 1–2%; ref. 10). The observation of preferential paternal but not maternal transmission of the diabetogenic allele *DR4* seems to explain half this effect.

Second, the study helps to elucidate the currently unexplained residue of miscarriages by identifying a genetic cause other than chromosomal structural abnormalities and blood-group incompatibility. Other genetic factors have also been noted recently¹¹.

Third, virtually all discussions of natural selection by evolutionary biologists have focused on postnatal advantages. But prenatal selection must also be enormously important, as 50–80% of human conceptuses, whose entire potential reproductive careers lie ahead of them, die before birth. The prenatal advantages of alleles deserve more attention; they could outweigh postnatal disadvantages of the same alleles.

Finally, the study of Vadheim *et al.* may resolve the evolutionary mystery of why insulin-dependent diabetes persists at all. It is true that only 20% of those children carrying the diabetogenic alleles actually develop the disease, but until recently those 20% would almost all have died without reproducing. Such massive removal of the alleles would reduce their frequency by three orders of magnitude within a millenium, unless they had some compensating advantage. Given a 20% postnatal disadvantage, the observed prenatal advantage (65–72 instead of 50% probability of inheritance) suffices to keep the alleles in circulation (ref. 3; see figure). Other recent evidence suggests that fetuses carrying a gene encoding cleft lip and palate also enjoy a prenatal advantage¹¹. Thus, insulin-dependent diabetes and cleft lip and palate may join sickle-cell anaemia and a growing number of other human genetic disorders recognized as being maintained by natural selection, as Jerome Rotter and I discussed¹² in our recent News and Views article. □

1. Sperling, K. in *Mutations in Man* (ed. Obe, G.) 128–146 (Springer, Berlin, 1984).
2. Lenton, E. A. *et al. Fertil. Steril. Suppl.* **47**, 18A (1987).
3. Vadheim, C. M. *et al. N. Engl. J. Med.* **315**, 1314–1318 (1986).
4. Vadheim, C. M. *et al. Diabetes* **34**, 21A (1985).
5. Vogel, F. & Motulsky, A. G. *Human Genetics* (Springer, Berlin, 1986).
6. Lauritsen, J. G. *Genetic Aspects of Spontaneous Abortion* (Aarhus University Press, 1977).
7. Martin, R. H. *et al. Hum. Genet.* **63**, 305–309 (1983).
8. Beer, A. E. *et al. Am. J. Obstet. Gynecol.* **141**, 987–999 (1981).
9. Reznikoff-Etievant, M. F. *et al. Tissue Antigens* **24**, 30–34 (1984).
10. Warram, J. H. *et al. N. Engl. J. Med.* **311**, 149–152 (1984).
11. Dronamraju, K. R. *et al. Clin. Genet.* **26**, 322–330 (1984).
12. Rotter, J. I. & Diamond, J. M. *Nature* **329**, 289–290 (1987).

Jared M. Diamond is Professor of Physiology at the University of California Medical School, Los Angeles, California 90024, USA.

Particle physics

Strings in four dimensions

John Ellis

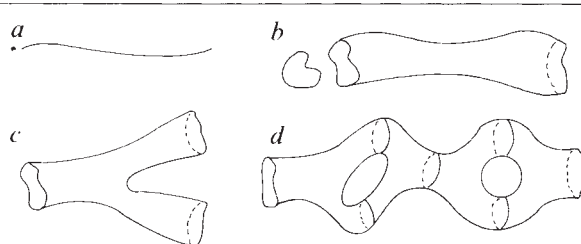
STRING theory provides for the first time a way to construct a consistent, finite theory of quantum gravity. It has a structure rich enough to contain all the other fundamental interactions and to explain many other basic features of physics, such as the variety of quark and lepton species. Small wonder, then, that many particle physicists believe that some variant of string theory is the 'theory of everything' or TOE. But which string theory? At first, consistency at the quantum level seemed to require that string theories be formulated in more than four dimensions, with the surplus ones compactified on a scale of the Planck length, 10^{-33} cm. Recently, however, string theorists have learnt how to

One of the many reasons for this was the discovery⁶ that the quantum theory of strings is consistent only in 26 dimensions — for bosons — or 10 dimensions for a supersymmetric theory including fermions. In spaces with different numbers of dimensions, the zero-point energy corresponding to quantum fluctuations of the string would be infinite, destroying not only the finiteness of the theory, but also an infinite-dimensional 'conformal' symmetry which was essential to remove states of negative probability and to maintain relativistic invariance.

This essential symmetry is generated by an infinite number of operators obeying algebraic relations discovered by Virasoro⁷. Conformal symmetry, the Virasoro algebra and its relatives are crucial in at least two other physical problems: phase transitions in two dimensions⁸; and the theory of solitons. For example, the different critical behaviours of two-dimensional phase transitions correspond to different representations of the Virasoro algebra⁷. Later, the importance of such representations for string theory will become clear.

The current explosion of interest in strings is stimulated by the possibility that they might provide a unified TOE¹⁰. Kaluza and Klein proposed in the 1920s that unified theories might require extra dimensions, so theorists were ready to postulate that the surplus dimensions required by string theory could be rolled up on a small scale, leaving four large space-time dimensions. The last technical obstruction to constructing a TOE was that of modular invariance, namely learning how to avoid infinities when summing over string configurations with non-trivial topology. This difficulty, phrased in different terms, was removed by M. B. Green and J. H. Schwarz¹¹, who left us with a finite number of meaningful and usable string theories in 26 and 10 dimensions.

Accordingly, much effort has been devoted to compactifying the most promising theory¹² to four dimensions and looking at the corresponding four dimensional model. This compactification involves mathematical gymnastics, and is ultimately unsatisfactory. The first attempts postulated compactification on a



a, A point particle (the dot) described by a field $\phi(x)$ propagates along a world-line, whereas *b*, a string (the loop) described by a non-local functional $\Phi(x)$ propagates over a world-sheet. Quantum fluctuations of the world-sheet have infinite zero-point energy unless the string lives in a space-time with a certain critical number of dimensions, or some of these are replaced by the appropriate number of internal degrees of freedom (conformal invariance). Interactions generate non-trivial topologies *c* and *d* whose consistent description also imposes important consistency conditions (modular invariance). Conformal and modular-invariant string theories can now be formulated directly in four dimensions.

formulate string theories directly in four dimensions¹⁴. Such theories can be described mathematically as representations of an infinite-dimensional algebra. They offer a wide new range of possibilities for phenomenological models, whose exploration is just beginning.

According to string theory⁵, elementary particles — including the photon, graviton, gauge bosons, quarks and leptons — are not point-like objects without structure, but are normal modes of oscillation of a string. A particle trajectory is a world-sheet in space-time, not a world-line as in the 'conventional' description of a point particle (*a* and *b* in the figure). Interactions are described by non-trivial topologies of the world-sheet, looking like trousers or the surfaces of Henry Moore sculptures (*c* and *d* in the figure). String theories were originally formulated as theories of the strong interactions, but were soon superseded in this role by quantum chromodynamics, the gauge theory of quarks and gluons.

smooth manifold¹³, many thousands of which are believed to exist. Subsequently it was realized that string theories could equally well be compactified on orbifolds¹⁴. These are mathematical spaces that are smooth everywhere apart from a discrete set of singularities.

More recently it has been realized that one could use different orbifolds for the modes of oscillation propagating clockwise and anticlockwise around the closed string¹⁵. These defy any simple interpretation in terms of compactification on some real space. The physical meaning of space dimensions no longer than the Planck length is questionable. It is clear that the process of compactification introduces enormous arbitrariness into what was originally almost a unique theory. One can avoid the problems of interpretation, and approach these ambiguities systematically by formulating a string theory in four dimensions.

The two key requirements for a four-dimensional string theory are conformal invariance, which involves the choice of a representation of the appropriate Virasoro algebra, and modular invariance, which is the consistent summation over non-trivial string configurations mentioned above. Conformal invariance is assured by replacing the surplus space dimensions with the correct number of internal degrees of freedom which cancel out the dangerous zero-point energy fluctuations.

These degrees of freedom could be combinations of bosons², fermions^{3,4}, or even general two-dimensional critical systems⁹ of the type I have mentioned above. Modular invariance imposes consistency conditions on the combination of these sub-theories, all of which are irreducible representations of the Virasoro algebra. Some of the degrees of freedom might have a space-time interpretation, but this is not necessary. Certainly, the previous attempts at compactifications of 10-dimensional string theories should be included among all these four-dimensional strings, but there are many other possibilities as well.

Not all the appropriate Virasoro representations are known, although it is known that there are many equivalences between different known representations — for example, between free bosons and pairs of free fermions, or between certain interacting fermion-fermion models and free fermions. The string model that yields convincingly the standard model of particle physics with three generations of quarks and leptons is yet to be found. Even when it is found, we must ask why our observable Universe chose this model rather than one of the multitude of other four-dimensional models, and even whether different choices are possible in different parts of the Universe.

More fundamentally, why did nature choose exactly four space-time dimen-

sions, and neither more nor less? Last, but not least, for string theory to be called physics, it must be susceptible to experimental test. Previously proposed signatures of strings turn out to be very model-dependent, and no-one has even described a smoking gun, let alone found one. String theory is still at the stage of travelling hopefully. □

1. Narain, K.S. *Phys. Lett.* **169B** 41–46 (1986).
2. Lerche, W., Lust, D. & Schellekens, A.N. *Nucl. Phys.* **B287**, 477–507 (1987).
3. Kawai, H., Lewellen, D.C. & Tye, S.H.H. *Phys. Rev. Lett.* **57**, 1832–1835 (1987).
4. Antoniadis, I., Bachas, C. & Kounnas, K. *Nucl. Phys.*

- B289**, 87–108 (1987).
5. Green, M.B. *Nature* **314**, 409–414 (1985).
6. Lovelace, C. *Phys. Lett.* **34B**, 500–506 (1971).
7. Virasoro, M. *Phys. Rev. D* **1**, 2933–2936 (1970).
8. Polyakov, A.M. *J. exp. theor. Phys. Lett.* **12**, 381–383 (1970).
9. Friedan, D., Ojii, Z. & Shenker, S. *Phys. Rev. Lett.* **52**, 1575–1578 (1984).
10. Ellis, J. *Nature* **323**, 595–598 (1986).
11. Green, M.B. & Schwarz, J.H. *Phys. Lett.* **149B**, 117–122 (1984).
12. Gross, D.J. *et al. Phys. Rev. Lett.* **54**, 502–505 (1985).
13. Candelas, P. *et al. Nucl. Phys.* **B258**, 46–74 (1985).
14. Dixon, L., Harvey, J.A., Vafa, C. & Witten, E. *Nucl. Phys.* **B261**, 678–686 (1985).
15. Narain, K., Sarmadi, M. & Vafa, C. Harvard preprint HUTP-86/AO89 (1986).

John Ellis is in the Theory Division, CERN, CH-1211 Geneva 23, Switzerland.

Transcription

At the heart of the nucleolus

Gwyn Jordan

Is it possible that nucleolin, the main protein of the nucleolus¹ of exponentially growing eukaryotic cells, regulates chromatin conformation, provides the basis for polymerase I specificity, controls formation of secondary structure in early ribosomal RNA transcripts and holds the whole transcription apparatus in an arrested state until it is phosphorylated and cleaved (Fig. 1)? Such an amazing set of activities could all belong to this

be the nucleolar analogue of an HMG, a small nuclear protein thought to evict histone (H1) from nucleosomes in the decondensation of chromatin (see Fig. 1). Nucleolin adds to chromatin at one molecule per nucleosome and has the same effect as removal of H1; some of the peptides of nucleolin bind to H1 (M. Erard, Toulouse). Olson and co-workers³ have already shown a binding preference of nucleolin for spacer sequences, the

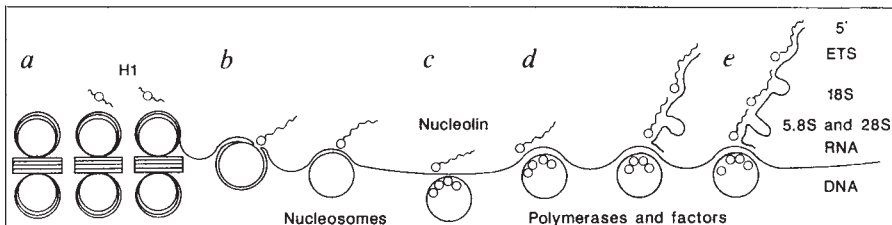


Fig. 1 In the nucleolus the genes that encode the RNA of ribosomes are transcribed. Nucleolin, a major nucleolar protein formerly called C23, affects the chromatin structure and RNA maturation of these genes. *a*, H1 coming off condensed chromatin; *b*, dispersed nucleolar chromatin with nucleolin bound instead of H1; *c*, pre-initiation complex on non-nucleosomal chromatin. The nucleosomes are lost at this stage, the large unit now represents polymerase I, and the small circles topoisomerase I and the three initiation factors; *d*, phosphorylation and proteolysis of nucleolin; *e*, transcription and formation of pre-ribosomal RNA secondary structure in preparation for processing. ETS, external transcribed spacer.

715-amino-acid protein, recently sequenced by F. Amalric and co-workers (Centre de Recherche de Biochimie et de Genetique Cellulaires du CNRS, Toulouse)². Nucleolin is the main silver-staining protein of nucleoli, and the elucidation of its amino-acid sequence provides some fascinating clues to its function, which were discussed at a recent workshop*.

There are three zones in the amino-acid sequence of nucleolin. First, the amino-terminal region contains six examples of the following sequences: GAV-Thr-Pro-GAV-Lys-Lys-GAV-GAV (where GAV is Gly, Ala or Val, that is, a non-polar residue) followed by three tracts of Glu and Asp residues at positions 141–167, 188–213 and 239–271. This zone may

regions lying between the repeated ribosomal RNA genes in eukaryotes. It is not clear how nucleolin recognizes nucleolar chromatin.

The second zone of nucleolin, near the carboxy terminus, has no hydrophobic regions and is thought to be in an extended conformation. It contains a sequence with 37 Gly, 5 Phe and 10 Arg residues, in which all the latter are dimethylated^{4,5}.

The third zone, the central part of the protein, which has alternating hydrophilic and hydrophobic segments, binds RNA² and has some similarities to the proteins of 'spliceosomes'. These properties probably account for nucleolin's demonstrated association with early transcribed RNA in the nucleolus⁶. B. Bugler (Amalric's group) reported on the sites of RNA binding in crosslinking experiments using

* 10th Nucleolar Workshop organized by W. Hennig, Stevens-beek, Nijmegen, The Netherlands, 27 June–1 July 1987.

ultraviolet light. With pre-ribosomes, a 40-nucleotide-long RNA sequence binds to nucleolin, and with the *in vitro* transcription system itself, binding sites appear on pre-ribosomal RNA near the 5' end as well as in the 18S and 28S regions.

It is tempting to speculate that nucleolin, or part of it (it is very easily fragmented), also regulates pol I transcription. (Pol I is the ribosomal RNA polymerase, pol II synthesizes messenger RNA and pol III transfer RNA and 5S RNA.) There is no direct evidence for this suggestion; in fact, nucleolin inhibits transcription until it is cleaved by a proteolytic step⁷. Before this proteolysis it is phosphorylated by the same kinase that phosphorylates topoisomerase I and the polymerase itself.

The idea that nucleolin could regulate transcription is consistent with other recent evidence concerning what happens to ribosomal RNA genes when no silver-binding proteins (of which nucleolin is one) can be detected on them at metaphase⁸. When ribosomal RNA genes, which are normally only ever transcribed by pol I, are transfected into the cells of another species and amplified by co-selection using the methotrexate system, they can be transcribed by pol II. Under these circumstances no silver staining is detected on the transfected group of genes in metaphase chromosomes. If this means that nucleolin is absent, rather than is present in concentrations too low to show up, nucleolin must be implicated in pol I selection. The explanation given in ref. 8 is that a species-specific factor controls polymerase selection and the human factor is absent from the hamster cells into which the human genes have been transferred. It is well known that species specificity exists for transcription of these genes. This has been analysed by M. Muramatsu and co-workers (University of Tokyo), among others, who used chimaeric promoter sequences constructed with a range of human-to-mouse ratios.

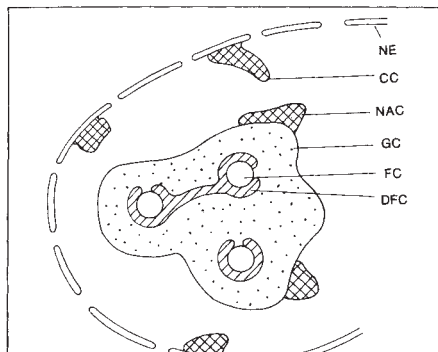


Fig. 2 The nucleolus is the site of ribosome synthesis. NE, nuclear envelope; CC, condensed chromatin; NAC, chromatin associated with the nucleolus; GC, granular component containing particles destined to become the large subunits of ribosomes; DFC, dense fibrillar component; and FC, fibrillar centre.

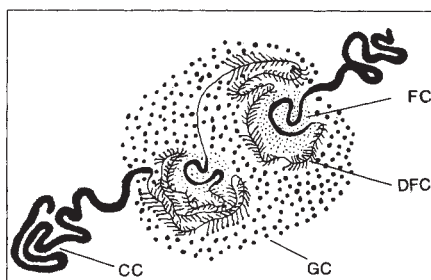


Fig. 3 Nucleolar structure. FC, pre-initiation complex of pre-ribosomal RNA genes, pol I, topoisomerase I and transcription initiation factors; DFC, active pre-ribosomal RNA transcription units; GC, pre-ribosomal particles, especially later stages of processing and protein addition to large ribosomal subunit.

These authors showed that the DNA sequence controlling specificity resides between residues -32 and -12 in mouse and includes the same region in humans but is extended further downstream there. The protein factor mediating this species specificity is one of several transcription initiation factors that have been described for pol I.

I. Grummt (Institute for Biochemistry, Würzburg) simplified the complexity that has grown up around the nomenclature of these factors by describing three identifiable functions known to be under their control: the first to act is the modulator; second, the species-specific function that sets up a stable initiation complex; and third, a factor controlling efficiency of initiation. These factors react with the promoter and the start of the genes in the region -40 to +23 (mouse) and also with the polymerase itself. Topoisomerase I, the DNA supercoil relaxer, is also required.

No one is yet prepared to call nucleolin, or part of it, a transcription factor. It is clear from the reported sequence that there are slight differences in the chromatin-binding domains from hamster and rat proteins, which seems significant in such a highly conserved protein. Is this difference related to the species-specificity function in transcription?

Where does it happen?

Fibrillar centres, small enigmatic structures deep in the nucleoli (Fig. 2), were ignored for years. Gradually, their importance has become apparent as it was realized that they contain the genes encoding ribosomal RNA. Their function and relation to the events of transcription were discussed at the workshop. Surprisingly, Scheer and Rose⁹ a few years ago demonstrated that fibrillar centres are the only places in nucleoli where pol I could be demonstrated in the electron microscope by immunogold labelling. All earlier electron-microscope autoradiographic evidence pointed to the surrounding dense fibrillar component as the site of

active transcription, and this component was therefore the expected location for pol I. It is not easy to dismiss Scheer's (University of Würzburg) findings, as he has now confirmed them with more elegant micrographs and shows that topoisomerase I is also in fibrillar centres. E. Egyhazi (Karolinska Institute, Sweden) and S.T. Jacob (Pennsylvania State University) presented evidence that pol I is associated with topoisomerase I during transcription. Together with R. Benavente (University of Würzburg), Scheer also shows that antibodies against pol I can stop transcription *in vivo* and *in vitro* and presumably are not therefore sterically hindered from getting to the active polymerases, further evidence that transcription takes place in the fibrillar centres.

O. L. Miller (University of Virginia) suggested that the fibrillar centre can be thought of as a chromomere from which the transcription units are reeled out. This is essentially the view of most microscopists and places the transcription events in the dense fibrillar component (Fig. 3). But what of the presence of pol I and topoisomerase I in the fibrillar centres? Although their function there may be debated, the evidence for their presence is undisputed. It is also worth recalling that nucleolin has been localized in both the fibrillar centres and the surrounding dense fibrillar component. The best answer at the moment seems to be that the fibrillar centre is the equivalent of a pre-initiation complex of ribosomal RNA genes. This means that at the heart of the nucleolus there are non-transcribing genes complexed with pol I, topoisomerase I and possibly the three transcription factors, but somehow held in a non-transcribing state, perhaps by nucleolin. On initiation of transcription, these genes transform themselves into the dense fibrillar component by the accumulation of nascent transcripts and ribosomal proteins. If their vacated position in the fibrillar centres is taken by more genes being prepared for transcription, we would have a dynamic cyclical process. The explanation for Scheer's results may partly relate to the speed of such events and the relative abundance of pol I and topoisomerase I in the respective components. □

1. Orriek, L.R., Olson, M.O.J. & Busch, H. *Proc. natn. Acad. Sci. U.S.A.* **70**, 1316-1320 (1973).
2. Lapeyre, B., Bourbon, H. & Amalric, F. *Proc. natn. Acad. Sci. U.S.A.* **84**, 1472-1476 (1987).
3. Olson, M.O.J. *et al. Biochemistry* **22**, 3345-3351 (1983).
4. Lapeyre, B. *et al. J. biol. Chem.* **261**, 9167-9173 (1986).
5. Lischwe, M.A., Cook, R.G., Ahn, Y.S., Yeoman, L.C. & Busch, H. *Biochemistry* **24**, 6025-6028 (1985).
6. Herrera, A.H. & Olson, M.O.J. *Biochemistry* **25**, 6258-6264 (1986).
7. Bouche, G., Caizergues-Ferrer, M., Bugler, B. & Amalric, F. *Nucleic Acids Res.* **12**, 3025-3035 (1984).
8. Dhar, V., Miller, D., Kalkarni, A. & Miller, O. *Molec. Cell Biol.* **7**, 1289-1292 (1987).
9. Scheer, U. & Rose, K.M. *Proc. natn. Acad. Sci. U.S.A.* **81**, 1431-1435 (1984).

Gwyn Jordan is in the Department of Biophysics, Cell and Molecular Biology, King's College London (KQC), London WC2B 5RL, UK.

Life (and death) in Lake Nyos

SIR—Walter Shearer¹ asked what effect the 1986 disaster that killed more than 1,700 people to the north of Lake Nyos, Cameroons, had had on life, particularly fish, in the lake. Shearer should be thanked for drawing this interesting question to our attention and for pointing out that there have been studies of the fish in other volcanic lakes in Cameroon. But the published reports of the British Scientific Mission² and the American group³, and unpublished Mission reports of the French, Italian or Japanese teams make no mention of life (or death) in the lake.

When I went to Cameroon, as part of the British Scientific Mission, in early September 1986, I was expecting to find that there had been a fish-kill in Lake Nyos and wondered what it would tell us about the composition of the toxic gas. At that time there were many theories in the air but little evidence concerning the nature of the event. However, when we got to Lake Nyos we found that there were no dead fish in the lake, no certain reports of dead fish and no canoes or fish traps near the lake. And local people who knew the area well were able to confirm that there had been no fish in the lake prior to the disaster. So the simple answer to Shearer's question "what happened to the fish that were living in Lake Nyos?", is that there were no fish. I would suspect that the general failure to mention the apparent absence of fish in Lake Nyos reflects the natural tendency of scientists to concentrate on positive rather than negative evidence. But negative evidence can sometimes be very important, and I think this may be just such a case.

Water from Lake Nyos is discharged over a natural weir which has a free fall of 22.3 m, so although there are fish in the streams and rivers near to the lake they have no direct access into the lake. But as Shearer has pointed out there are endemic species of cichlid fish in crater lakes, access to which must also be restricted (otherwise the species would not be endemic). To transfer a breeding population of cichlid fish into an isolated lake, by natural means, presents less of a problem than it would for most other groups of fish as cichlids are mouth brooders. So it would need only one brooding parent to be taken by a bird of prey and dropped, dead or alive, into an otherwise isolated lake.

If lake Nyos were extremely young that might offer a possible explanation for the absence of fish. But Lake Nyos is, in my opinion, of a similar age to the other isolated volcanic lakes in which there are a number of species of fish.

There is no reason to suspect that under normal circumstances the waters of Lake Nyos would be hostile to naturally intro-

duced fish. From the limited amount of chemical data available (mainly in unpublished mission reports) we can be reasonably sure that prior to the disaster the surface water was well oxygenated, mildly acid, and did not contain high concentrations of potentially toxic ions.

The available evidence, some of which is outlined above, leads me to conclude that breeding populations of fish must surely have been introduced into Lake Nyos on a number of occasions. But that whenever this has happened a subsequent lake overturn, similar to that which caused the August 1986 disaster, completely deoxygenated the surface water and killed off all the fish. As there were no fish in the lake before the recent disaster the probability must be that major gas releases are rather more frequent than the natural introduction of viable fish populations—I would speculate that major gas releases might occur every 50 years or so.

Although we did not find any fish in Lake Nyos that does not mean to say that the lake waters were completely lifeless at the time of our visit. Whilst examining the south-western margin of the lake I came across an apparently healthy population of frogs living by the side of the lake and hopping in and out of the water without apparent ill effect. If any expert on West African frogs or on the tolerance of frogs in general to toxic gases should happen to read these comments I would be very pleased to hear from them.

S. J. FREETH

Department of Earth Sciences,
University College of Swansea,
Swansea SA2 8PP, UK

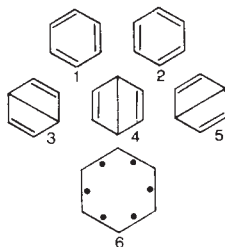
1. Shearer, W. *Nature* **328**, 299–300 (1987).

2. Freeth, S.J. & Kay, R.L.F. *Nature* **325**, 104–105 (1987).

3. Kling, G.W. *et al. Science* **236**, 169–175 (1987).

The electronic structure of the benzene molecule

SIR—I wish to comment on the interesting conclusions^{1–9} obtained from recent *ab initio* calculations for C₆H₆ (refs 1, 8) and to relate them to the earlier work of Linnett, Cooper, Gerratt and Raimondi¹ show that resonance between the Kekulé and Dewar valence bond (VB) structures (1)–(5) generates a lower energy than does the self consistent field (SCF) molecular orbital (MO) wave function and therefore,



to a good approximation, there is no delocalization of the six π electrons.

In 1965, Empedocles and Linnett¹⁰

calculated that the π electron energy for resonance between the Kekulé structures, with the Heitler–London formulation for the π electron wave functions, also lies below the SCF-MO energy for these electrons. No spatial optimization of the atomic orbitals (AOs), that is, no spreading of each AO towards its neighbour², was involved.

Linnett and his coworkers (ref. 11 and refs therein) obtained a similar result for resonance between two covalent structures of some systems that involve three-electron and three-overlapping AO bonding units—for example, H₃ (three σ electrons) and C₃H₅, NO₂, HCO₂ and O₃⁺ (each with three π electrons in the calculations). The covalent structures for each of these systems (for example H—H—H and H—H—H for H₃) correspond to the Kekulé structures for C₆H₆; each overlapping AO is singly-occupied and bonding occurs between pairs of adjacent atoms.

Lower energies were, of course, obtained when the one-centre AOs of the Kekulé-like structures were replaced by spatially-optimized orbitals¹² in the Heitler–London wave functions for these systems.¹¹ Even lower energies were calculated^{10,11} for Linnett's non-paired spatial orbital (npso) structures such as (6) for C₆H₆, in which each of the π electrons is located in a separate two-centre bonding MO. Thus, as Linnett often stressed¹³, the npso structure for C₆H₆, with localized (one-electron) π bonds, provides a better VB representation than does resonance between the Kekulé structures.

Messmer and Schultz⁸ have also calculated that resonance between the Kekulé structures gives a lower energy than does the SCF-MO configuration; they have also provided theoretical evidence that the double bonds of the Kekulé structures may be bent rather than σ and π . Bent bond descriptions¹³ also obtain for twelve of the C—C bonding electrons in Linnett's double quartet representation of the npso structure (6).

As indicated by Cooper, Gerratt and Raimondi, interconversion between the Kekulé structures (1) and (2) may be achieved by changing the spin-pairings. For electron-rich systems with four-electrons and three overlapping AO bonding units, such as H₃[−], O₃ and C₃H₅[−], electron delocalization is required in order to convert one Kekulé-like structure (for example, H[−]—H—H for H₃[−]) into the other (H—H—H[−]). When Dewar-like structures with long or formal bonds (for example, H—H—H) are also included in the resonance scheme, resonance has been calculated to generate a substantially lower (four-electron) energy than does the corresponding SCF-MO wave function^{14–16}. The relevance of this type of result for qualitative descriptions of bonding has been discussed¹⁷, primarily via

'increased-valence' structures.

Linnett's computational procedures could not match those now used, but because the npso structures involve better electron charge correlation than do the Kekulé structures^{10,13}, it is likely that Linnett's conclusion concerning the npso structure for C₆H₆ will be found to be valid when modern computational procedures are employed.

RICHARD D. HARCOURT

Department of Physical Chemistry,
University of Melbourne,
Parkville, Victoria, 3052, Australia

- Cooper, D. L., Gerratt, J. & Raimondi, M. *Nature* **323**, 699-701 (1986).
- McWeeny, R. *Nature* **323**, 666-667 (1986).
- Gerratt, J. *Chem. Brit.* **23**, 327-330 (1987).
- Emsley, J. *New Scientist* **112**, 25 (1986).
- Chem. Engng. News* **64**, 24 (1986).
- Pauling, L. *Nature* **325**, 396 (1987).
- Berry, R. W. H. *Chem. Brit.* **23**, 210 (1987).
- Messmer, R. P. & Schultz, P. A. *Phys. Rev. Lett.* **58**, 2416-2419 (1987).
- Maddox, J. *Nature* **327**, 551 (1987).
- Empedocles, P. B. & Linnett, J. W. *Proc. R. Soc. A* **282**, 166-177 (1965).
- Nibler, J. W. & Linnett, J. W. *Trans. Faraday Soc.* **64**, 1153-1164 (1968).
- Coulson, C. A. & Fischer, I. *Phil. Mag.* **40**, 386-393 (1949).
- Linnett, J. W. *The Electronic Structures of Molecules* (Methuen, 1964).
- Bowen, H. C. & Linnett, J. W. *Trans. Faraday Soc.* **60**, 1185-1192 (1964).
- Harcourt, R. D. & Harcourt, A. J. *chem. Soc. Faraday* **2**, 70, 743-757 (1974).
- Hiberty, P. C. *Israel J. Chem.* **23**, 10-20 (1983).
- Harcourt, R. D. *Lecture Notes in Chemistry* **30** (Springer, Heidelberg, 1982).

SIR—We feel it necessary to correct some misconceptions that two articles by John Maddox (*Nature* **324**, 599; 1986 and **327**, 551; 1987) on our recent work^{1,2} and that of Cooper *et al.*³ may have engendered.

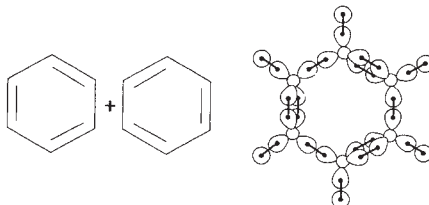
When firsts are first. The original article (about our work¹ on C₂F₂) was entitled "When firsts are really second" and took issue with the use of the word "first" to describe the results of our work. Specifically, the article suggested that our objective was to show that the "valence bond description may be a more accurate starting point for calculation of electronic structure of molecules than the molecular orbital framework" and that this conclusion had previously been made a month earlier in *Nature* by Cooper *et al.*³, refuting our claim to precedence.

What we actually had claimed was to have provided quantitative justification for the equivalent bent-bond representation (as opposed to the σ and π bond representation) of the triple bond, a description first proposed by Slater⁴ and Pauling⁵ long ago. The gross misstatement of our intent (in *Nature*) is extremely disconcerting and somewhat mystifying considering that our true objective is stated explicitly in the abstract: "These represent the first quantitative calculations which document a case where a triple 'bent-bond' description is energetically favoured over a σ, π bond description".

Bent bonds are better. The fundamental thrust of our work^{1,2} has been to address an old, but, until recently, unresolved

question: are multiple bonds in molecules better represented in terms of σ and π bonds or in terms of Ω bonds (equivalent multiple bent bonds)? Within valence bond theory (which is a higher level of theory than that of molecular orbitals), this question can be addressed. (The two descriptions are energetically indistinguishable in molecular orbital theory.)

Our work on C₂F₂ demonstrated that triple bonds are energetically better represented in terms of Ω bonds. Earlier work had established the Ω bond description for double bonds^{6,7} and later work proved this also to be the case for the canonical



σ, π system, benzene². Based on the variational principle (energetic criterion), we found that the benzene molecule was best represented in terms of two Kekulé structures where each structure was described in terms of alternating single and equivalent bent double bonds about the ring as shown in the figure. These calculations for a series of molecules represent the first quantitative evidence of the superiority of the Ω bond description, which has caused a reexamination of the conventional views of bonding. In their calculations on benzene, Cooper *et al.*³ do not even mention this most fundamental point.

When firsts are really second. The main theme of the second article was the desirability of adequate citation of complementary work. We are in full agreement that it is proper to cite work that serves as a foundation to one's own contribution. We believe, however, that the conclusion erroneously attributed to us (see first point above), and seemingly discovered by Cooper *et al.*³ is obvious and certainly predates both publications. It should be pointed out that calculations using the valence bond approach date from 1927, with the calculation of Heitler and London on the H₂ molecule⁸, and that valence bond theory was given its most powerful articulation by Pauling⁹.

Our own calculations are based on a theory first given formal expression by Hurley *et al.*¹⁰ and eventually implemented as a special case of the generalized valence bond theory¹¹⁻¹³. It would be futile to attempt to catalogue all the valence bond calculations carried out since then. Nor are calculations on benzene novel. A mention of the calculation of Cooper *et al.*³ would have added little to our discussion. An *ab initio* valence bond calculation on benzene was reported in 1973 (ref. 14), a self consistent version in 1974 (ref.

15). Indeed, Pauling has noted¹⁶ that he had predicted the results of Cooper *et al.* in 1933. The high quality configuration interaction calculations of Hay and Shavitt¹⁷ and the localized molecular orbital study of Newton and Switkes¹⁸ were pertinent to our discussion as they question the σ, π orbital description.

Prudence. Hay and Shavitt¹⁷ reported a self consistent field (SCF) energy of -230.6410 hartree. Using the same gaussian basis set (with a minor difference in the contraction), Cooper *et al.*³ calculate an SCF energy -229.9954, which is more than 17 eV higher. Our calculation², again with the same basis set but this time with the same contraction as Cooper *et al.* finds an SCF energy of -230.6404, in agreement with Hay and Shavitt¹⁷. The magnitude of the discrepancy, 17.6 eV, points to only one conclusion: the calculation of Cooper *et al.* is flawed.

As the SCF value had been published over ten years earlier, the lapse in responsibility of the authors to examine the literature and the failure of the system of review to catch this oversight can only serve as an embarrassment to both authors and editor. Thus we believe your critical comments to be without foundation and take exception to them. The interested reader should also note the article in *New Scientist* **115**, 30 (1987).

RICHARD P. MESSMER*†

PETER A. SCHULTZ*

* Department of Physics and Laboratory for
Research on the Structure of Matter,
University of Pennsylvania,
Philadelphia, Pennsylvania 19104, USA

† General Electric Corporate Research and
Development,
Schenectady,
New York 12301, USA

- Messmer, R. P. & Schultz, P. A. *Phys. Rev. Lett.* **57**, 2653 (1986).
- Schultz, P. A. & Messmer, R. P. *Phys. Rev. Lett.* **58**, 2416 (1987).
- Cooper, D. L., Gerratt, J. & Raimondi, M. *Nature* **323**, 699 (1986).
- Slater, J. C. *Phys. Rev.* **37**, 481 (1931).
- Pauling, L. *J. Am. chem. Soc.* **53**, 1367 (1931).
- Messmer, R. P., Schultz, P. A., Tatar, R. G. & Freund, H.-J. *Chem. Phys. Lett.* **126**, 176 (1986).
- Palke, W. E. *J. Am. chem. Soc.* **108**, 6543 (1986).
- Heitler, W. & London, F. Z. *Phys.* **44**, 455 (1927).
- Pauling, L. *The Nature of the Chemical Bond*, 3rd edn (Cornell University Press, Ithaca, 1960).
- Hurley, A. C., Lennard-Jones, J. E. & Pople, J. A. *Proc. R. Soc. A* **220**, 446 (1953).
- Hunt, W. J., Hay, P. J. & Goddard, W. A. III *J. chem. Phys.* **57**, 738 (1972).
- Goddard, W. A. III, Dunning, T. H. Jr, Hunt, W. J. & Hay, P. J. *Acc. Chem. Res.* **6**, 368 (1973).
- Bobrowicz, F. W. & Goddard, W. A. III, in *Modern Theoretical Chemistry* Vol. 3 (ed. Schoefer, H. F. III) 79 (Plenum, New York, 1977).
- Norbeck, J. M. & Gallup, G. A. *J. Am. chem. Soc.* **95**, 4460 (1973); **96**, 3386 (1974).
- Levin, G. B. thesis, Caltech (1974).
- Pauling, L. *Nature* **325**, 396 (1987).
- Hay, P. J. & Shavitt, I. *Chem. Phys. Lett.* **22**, 33 (1972); *J. chem. Phys.* **60**, 2865 (1974).
- Newton, M. D. & Switkes, E. *J. chem. Phys.* **54**, 3179 (1971).

SIR—John Maddox (*Nature* **327**, 551; 1987) has discussed two calculations of the electronic structure of the benzene

molecule, one by Schultz and Messmer¹, the other by ourselves². We wish to clarify several important issues.

In our spin-coupled valence bond (VB) calculation on benzene², the first step is the separation of the σ and π electrons. The 36σ electrons are accommodated in a core of doubly filled self consistent field (SCF) orbitals and only the six π electrons are explicitly treated. No further preconceptions are imposed.

The orbitals and the associated coupling of the electron spins are fully optimized so as to minimize the total energy. The orbitals may be localized or delocalized but, as it turns out, the most stable solution is one in which the orbitals are in fact localized. Similarly the type of spin coupling is not imposed. The preponderance of the two Kekulé structures and the much smaller role played by the para-bond (or Dewar) structures (see (3), (4) and (5) in the preceding letter) emerges simply as a result of the calculation. Refinement of the wave function by means of non-orthogonal configuration interaction does not change the physical picture, which consequently carries conviction.

The work of Schultz and Messmer¹ uses a slight modification of the "separated electron-pair wave function" of Hurley, Lennard-Jones and Pople^{3,4}, now more commonly known as the perfect pairing 'GVB' wave function. This approach involves strong orthogonality constraints in which the orbitals are placed in pairs. Each orbital is allowed to overlap with its partner, but not with orbitals from other pairs. Furthermore, only one coupling of the spins is allowed, which amounts to constraining the electrons in each pair to have opposing spins. Thus in a GVB calculation on benzene, a Kekulé structure is chosen and associated with one ordering of the orbitals. In order to preserve the overall symmetry of benzene, Schultz and Messmer have taken the sum of two such structures (or spin couplings).

It is clear from this that the GVB calculation incorporates many preconceptions. In the particular case of benzene, where there are no obviously directed bonds, the imposition of the strong orthogonality constraint is quite inappropriate; in particular, it is not obvious which overlaps can be taken to be zero. We believe that, as a consequence of the restrictions inherent in this approach, localized 'banana' bonds of the type found are to all intents and purposes guaranteed from the outset. If there had been no σ - π mixing, then there would instead simply be three localized π bonds in each of the Kekulé structures.

Recent spin-coupled calculations on the HCCH and N₂ molecules have involved ten non-orthogonal orbitals and five spin functions, all of which were fully optimized. The spin-coupled calculations involve the largest number of non-

orthogonal orbitals ever used in such a calculation, while the number of spin functions is sufficient to allow for the uncoupling which occurs when the triple bond is broken. (The results will be published in due course.) No separation of σ and π orbitals was assumed, but preliminary results show that the orbitals spontaneously separate into functions of pure σ and pure π symmetry. There is no sign of 'bent' or 'banana' bonds, which must therefore correspond to higher energy.

In general, the application of the strong orthogonality constraint needs much care, and indiscriminate use of the GVB method can lead to seriously misleading results. This point is well illustrated by diazomethane (CH₂N₂), which belongs to a large class of molecules known as 1,3-dipoles. The GVB calculations on such systems indicate that they are essentially singlet diradicals—for which there is no available experimental evidence. On the other hand, spin-coupled calculations show quite unambiguously that these molecules have very little diradical character⁵. In spite of some very large overlaps between orbitals, which makes any kind of orthogonality constraints unrealistic, the electrons pair up to form reasonably normal bonds.

It is pleasing that the calculations of Schultz and Messmer lend some support to our view that the orbitals on benzene localize on account of electron correlation, but the evidence suggests very strongly that the occurrence of bent bonds is an artefact arising because of the constraints imposed upon the wave function, and has no physical significance.

J. GERRATT

School of Chemistry,
University of Bristol,
Bristol BS8 1TS, UK

M. RAIMONDI

Department of Physical Chemistry
and Electrochemistry,
University of Milan,
Via Golgi 19,
20133 Milan, Italy

D. L. COOPER

Department of Physical, Inorganic
and Industrial Chemistry,
University of Liverpool,
PO Box 147,
Liverpool L69 3BX, UK

- Schultz, P. A. & Messmer, R. P. *Phys. Rev. Lett.* **58**, 2416 (1987).
- Cooper, D. L., Gerratt, J. & Raimondi, M. *Nature* **323**, 699 (1986).
- Hurley, A. C., Lennard-Jones, J. & Pople, J. A. *Proc. R. Soc. A220*, 446 (1953).
- Gerratt, J. & Lipscomb, W. N. *Proc. natn. Acad. Sci. U.S.A.* **59**, 332 (1968).
- Cooper, D. L., Gerratt, J., Raimondi, M. & Wright, S. C. *Chem. Phys. Lett.* (in the press).

MHC antigen expression in the pancreas

SIR—Pujol-Borrell and his colleagues report (*Nature*, **326**, 304–306; 1987) that the combined action of interferon γ (IFN- γ) and tumour necrosis factor (TNF) or lymphotoxin is required to induce expression of class II major histocompatibility complex (MHC) antigens on pancreatic islet cells containing insulin, glucagon, and somatostatin and conclude that this observation "not only strengthens the previous histopathological observations of the pancreas of diabetic individuals but also points to a possible sequence of events leading to this phenomenon *in vivo*". They have previously proposed that aberrant expression of class II MHC antigen on endocrine cells may be an initiating feature of autoimmune disease¹.

It is difficult to understand how these authors can use their current data either to strengthen previously reported observations of pancreatic histopathology in Type I diabetes both by themselves and others because these have invariably shown selective expression of class II MHC antigen restricted to islet cells containing insulin in the pancreas of both patients^{2,3} and animals⁴, or to support their concept that expression of class II MHC antigen plays a key role in initiating the autoimmune destruction of the insulin-secreting cells of the pancreas.

In contrast, our previously reported *in vitro* studies⁵ using isolated pancreatic islets obtained from BB rats (which after a prediabetic period spontaneously develop autoimmune, insulin-dependent diabetes⁶) are consistent both with the pancreatic histopathology seen in Type I diabetes and with our concept of the sequence of events leading to the development of overt insulin-dependent diabetes⁴.

Thus we have shown that IFN- γ alone induces expression of class II MHC antigens only on insulin-containing cells in islets derived from prediabetic BB rats and have never observed induction of class II MHC antigen expression by IFN- γ on (1) cells containing glucagon or somatostatin in islets obtained from prediabetic rats, (2) any endocrine cells in islets obtained from normal Wistar rats, rats from a diabetes-resistant BB line, or (3) very young (less than 30 days of age) prediabetic BB rats. We postulated that the insulin-containing cells in diabetes-prone BB rats become specifically sensitized to IFN- γ in the prediabetic period, perhaps by cytokine-inflicted damage, and, like Pujol-Borrell and his colleagues, concluded that more than one agent is required for the induction of class II MHC antigen expression on pancreatic insulin-containing cells in diabetes-prone animals. However, the fact that IFN- γ is produced only by T cells suggested to us that this expression of class II antigen was more likely to be a consequence than the

cause of the mononuclear cell infiltration of the pancreas seen in Type I diabetes.

J.D. BAIRD
A.J. BONE
R. WALKER

Metabolic Unit,
University Department of Medicine,
Western General Hospital,
Edinburgh EH4 2XU, UK

A. COOKE

Department of Immunology,
Middlesex Hospital Medical School,
London, UK

1. Bottazzo, G.F., Pujol-Borrell, R., Hanafusa, T. & Feldman, M. *Lancet* **ii**, 1115-1119 (1983).
2. Bottazzo, G.F. *et al.* *New Engl. J. Med.* **313**, 353-360 (1985).
3. Foulis, A.K. & Farquharson, M.A. *Diabetes*, **35**, 1215-1224 (1986).
4. Dean, B.M., Walker, R., Bone, A.J., Barid, J.D. & Cooke, A. *Diabetologia*, **28**, 464-466 (1985).
5. Walker, R. *et al.* *Diabetologia*, **29**, 749-751 (1986).
6. *Metabolism* **32**, suppl. 1 (1983).

AIDS incubation period

SIR—Jonas Salk (*Nature* **327**, 475-476, 1987) postulates that the long incubation period between infection and development of clinical AIDS (acquired immune deficiency syndrome) may be due to an immune response to the initial infection which, if boosted, would reduce the viral burden and prevent development of the clinical spectrum of AIDS.

An essential part of the hypothesis is that loss of CD4 cells is a consequence of both immune destruction of these cells by the host and by virus-induced cytopathicity; uninfected cells are coated with viral protein products and, as a result, are destroyed by antibody-dependent cell cytotoxicity, leading to a net loss of CD4 lymphocytes. The destruction of uninfected CD4 cells in turn diminishes antibody-dependent cell cytotoxicity.

Professor Salk might have mentioned in support of his argument the known clinical situation in which repeated immune modulation continues after exposure to human immunodeficiency virus (HIV) infection. The striking example of this is haemophilia. Patients with severe haemophilia average at least one infusion of factor VIII concentration a week which contains not only factor VIII and variable amounts of other non-autologous proteins but also significant quantities of immunoglobulins. The rate of conversion of HIV infection to clinical AIDS in haemophilia is about 3-5 per cent of the infected population, much less than that in any other risk group. So far, there has been no adequate explanation of this ten fold factor. One possibility is that, in haemophilia, the incubation period will be longer than in other infected groups and this is supported by some mathematical models which predict incubation periods of five to fifteen years following blood-borne infection. Another possibility, however, is that repeated injections of factor VIII concentrates do in fact alter the immune response

of haemophiliacs so that HIV protein products which attach to uninfected CD4 cells are more quickly cleared and/or that antibody-dependent cytotoxicity to HIV-infected CD4 cells is enhanced.

Whatever the mechanism, Salk's hypothesis is an attractive way of explaining the observed low incidence of AIDS in haemophiliacs; the availability of blood specimens from such patients makes it possible to test his hypothesis.

H. EKERT

Department of Clinical Haematology
and Oncology,
Royal Children's Hospital,
Parkville, Victoria, 3052 Australia

Processing or inactivation of antimicrobial peptides

SIR—Three years ago while studying the skin secretions of *Xenopus laevis* we isolated and sequenced some of the peptides released on adrenaline stimulation. One was a undecapeptide with a sequence (see figure) that did not correspond to any known peptide or precursor at that time. The sequence remained unnoticed on the bulletin board in the laboratory until a recent News and Views column¹ disclosed our lack of imagination and the fact that it corresponds to an amino-terminal fragment of magainin II². The large amounts of this small molecular form of magainin II in the skin secretion could be a result either of biosynthetic processing or of post-secretory proteolytic degradation.

Part of the normal biosynthetic process (in toad skin glands also³) is the removal by a carboxypeptidase of the basic residues that have served their purpose as cleavage signals. These basic residues, however, are still found at the carboxyterminus of the magainin fragment. Thus, our undecapeptide fragment of the 23 amino-acid magainin II is probably a post-secretory degradation product.

This kind of degradation may serve as a protective inactivation mechanism occurring on the surface of the cells of the toad. The antimicrobial effect of the magainins

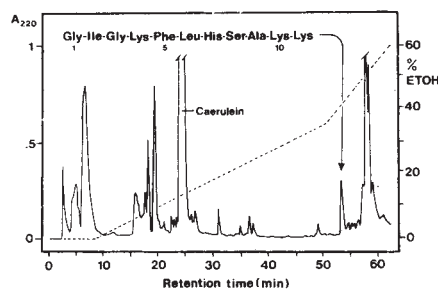


Fig. 1 Reconstituted ethanol washing of the skin of a toad injected subcutaneously with 50 µg of adrenaline, injected on a Vydac C₁₈ column. (1 × 25 cm) and eluted with TFA/H₂O (0.1%) and a gradient of ethanol. The peptide indicated by the arrow, which was sequenced on an Applied Biosystems 470A sequencer, corresponds to residues 1-11 of magainin II².

has been connected to their ability to form amphiphilic alpha-helices which can perturb membranes¹. Cleavage of the magainins (magainin I also has a tryptic-like cleavage site in the middle of the molecule) results in the generation of two peptides both too small to span the membrane as an alpha-helix. The basic antimicrobial peptides of insects, the cecropins⁴, can be cleaved and possibly inactivated by a similar mechanism.

THUE W. SCHWARTZ
H. FREDE HANSEN
ULRIK GETHER
LARS AAKERLUND

MRC Laboratory of
Molecular Endocrinology,
Rigshospitalet 6321,
DK-2100 Copenhagen, Denmark

1. Cannon, M. *Nature* **328**, 478 (1987).
2. Zasloff, M. *Proc. natn. Acad. Sci. U.S.A.* **84**, 5443-5449 (1987).
3. Richter, K., Egger, R. & Kreil, G. *J. biol. Chem.* **261**, 3676-3680 (1986).
4. von Hofsten, P. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **82**, 2240-2243 (1985).

Antimicrobial peptides

SIR—Following my News and Views article¹ on the antimicrobial magainin peptides of Zasloff^{2,3} I have become aware of a relevant publication of Giovannini *et al.*⁴ which was published in April and which I failed to cite in my text. I apologize to the authors for this omission.

Giovannini *et al.*⁴ report the detection and analysis of a wide range of peptides using skin secretions of *Xenopus laevis*. Two of these peptides, designated PGS and [Gly¹⁰,Lys²⁷]PGS have identical amino-acid sequences to, respectively, the magainins II and I described by Zasloff and co-workers^{2,3}. It is noteworthy, however, that Zasloff² failed to detect the magainins in mucous-rich secretions from *X.laevis* skin.

Precise physiological roles for biologically-active peptides are hard to pin down¹. Giovannini *et al.*⁴ suggest granular gland secretions have a defensive role as have others before them — for example, Richter *et al.*⁵, who also speculated more widely and considered a possible antibacterial activity associated with the need to combat infections that might befall *X.laevis* in its aqueous environment. Clearly, more experimental evidence is required before any definite conclusions can be reached concerning the functions of these peptides.

MICHAEL CANNON

Department of Biochemistry,
King's College London,
Strand, London, WC2R 2LS, UK

1. Cannon, M. *Nature* **328**, 478 (1987).
2. Zasloff, M. *Proc. natn. Acad. Sci. U.S.A.* **84**, 5443-5449 (1987).
3. Zasloff, M., Martin B. & Chen, H-C. *Proc. natn. Acad. Sci. U.S.A.* (in the press).
4. Giovannini, M.G., Poulter, L., Gibson, B.W. & Williams, D.H. *Biochem. J.* **243**, 113-120 (1987).
5. Richter, K., Egger, R. & Kreil, G. *J. biol. Chem.* **261**, 3676-3680 (1986).

Trailing in research work

Philip Gummert

Knowing Everything About Nothing: Specialization and Change in Scientific Careers.

By John Ziman. Cambridge University Press: 1987. Pp. 196. £15, \$29.95.

THESE are turbulent times for science in many countries. Rising costs and greater governmental concern about industrial competitiveness are leading to more pressure for exploitability, selectivity and concentration, all within a context of top-down management.

Nowhere is this more true than in Britain, where one effect of this more purposive approach to the management of science is that lines of research are more likely to be stopped than in the past. Partly for this reason, and partly because of the increasing tendency towards short-term funding of posts (compounded, we are told to expect, by the abolition of tenure in British universities), people who wish to make their career in science must increasingly expect from time to time either to be pressed to change their field of research, or actively seek change themselves in order to stay in fields that are likely to gain support.

But, as John Ziman reminds us, whereas it is said of philosophers that they learn less and less about more and more, until they know nothing about everything, of scientists it is said that they learn more and more about less and less, until they know everything about nothing. In science, specialization is the norm. Yet having spent years acquiring the knowledge and craft skills of a particular specialism, how possible is it for a scientist, at the drop of a governmental axe, to switch to a new specialism?

It is this question, posed against the backdrop of the effects of British government policy in recent years, that Ziman examines. Although the book is based on

British experience, the issues raised are of much wider interest.

The specific stimulus to the book was a meeting with the former Secretary of the Natural Environment Research Council (NERC), Mr (now Professor) R.J.H. Beverton. NERC had had radically to reorganize and re-orientate some of its research establishments, and found that a number of scientists had strongly resisted attempts to move them to different research areas, even when they seemed well qualified to take up the new field.

Ziman and Beverton decided to explore in detail the sources of resistance of this kind. Was it caused by the work that the scientists were doing, or by some aspect of their professional careers, or by the way that they were employed? Was there a stage in their careers when they became so unadaptable that it was scarcely worth the managerial effort to redeploy them on new problems? It soon became clear that, although a conventional wisdom could be said to exist on these questions, very little serious research had been done on them. Sociologists of science have studied research specialities, but not scientists as specialists. The literature is unhelpfully silent on what is now a pressing question in science policy and the management of research and development.

What Ziman and Beverton therefore did, after preliminary discussions with some heads of research councils, departmental chief scientists, directors of research laboratories, and similar people in several countries, was to engage in discussions with groups of staff at 15 research establishments. Seven of these were run

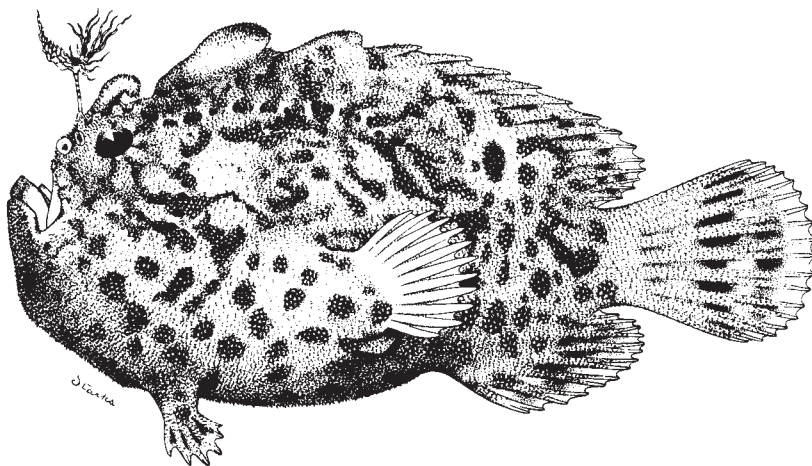
by research councils, five were in the public sector, two were in private industry and one was in an academic institution. The book is largely based upon these discussions. Indeed, it has more the flavour of a set of well-honed research notes than of an orthodox monograph. This is not to say that the book is difficult to follow: on the contrary, it has all the lucidity that is normally associated with Ziman's work. The point, rather, is that instead of packaging what the scientists had to say in the language of sociology or management of research and development, Ziman tries to let his readers in on the conversations themselves. We are invited to listen to what the scientists actually had to say, by means of lengthy quotations from the meetings.

The resulting news is both good and bad. Ziman shows how scientists typically spend their careers following what he calls research trails, which are the accumulation of problems with which a scientist engages during his or her career. Most scientists' research trails remain for long periods within one speciality, changing only gradually in terms either of topic or of techniques used. Substantial change is not normally sought.

Nevertheless, a noticeable proportion of scientists follow diverse careers, moving between a wide variety of specialities. Moreover, most scientists are required to learn new skills throughout their career, either because the emphasis of their field shifts, or because of changes in instrumentation, such as have most notably been occasioned by the rise of microcomputers. Typically, scientists in any case use in their work a variety of skills, many of which are drawn from other fields and therefore could be said to equip them for movement. Scientists are also well used to teaching themselves what they need to know about related fields.

What distinguishes the scientist who freely changes fields from the one who sticks to his or her last? No single answer is

Gone fishing — the bloody frogfish Antennarius sanguineus Gill, so-called because of its characteristic reddish-brown to dark-brown spots. The frogfishes (family Antennariidae, one of 18 families of anglerfishes) are distinguished by their unusual frog-like appearance and tetrapod-like locomotion. Brilliantly coloured but cunningly camouflaged to resemble sponge- or coral-encrusted rocks, they attract their prey by wiggling a highly conspicuous lure, a uniquely modified dorsal fin. Frogfishes are the fastest known predators, taking just 6 milliseconds to gulp down their prey. The illustration is taken from Frogfishes of the World: Systematics, Zoogeography, and Behavioral Ecology by Theodore W. Pietsch and David B. Grobecker, just published by Stanford University Press. Price of the book is \$67.50.



offered by Ziman, but a number of possibilities are raised, not the least of which is motivational factors. On the basis of his interviews, Ziman rejects the view that staleness and rigidity necessarily accompany increasing age (a point that would, I suspect, be confirmed by a study of the careers of many of the British academics who have taken early retirement since the university cuts of 1981). Rather, what matters is whether or not commitment is maintained, and this, of course in large measure returns the problem of flexibility to the realm of management.

Ziman's analysis is open to the criticism that it focuses too much on research council and government scientists and not enough on industrial and university scientists. The differences that might exist between these sectors in the management of scientists' careers might well repay analysis — the more so as, for example, industrial and civil service practices, such

as staff appraisal, are being introduced into British universities, but without the associated career development opportunities that are to be found in these other sectors.

There is, however, a limit to the fieldwork that is possible in any piece of research. As it is, we have here a workmanlike study that opens up an important field and raises numerous hypotheses for further examination. The crucial question, on which more data are likely to become available in the immediate future, is whether the patterns of behaviour that characterize the voluntary movement of scientists between fields can be transferred to circumstances in which there is more pressure for involuntary change. □

Philip Gummert is in the Department of Science and Technology Policy, University of Manchester, Manchester M13 9PL, UK. Among other books he is author of Scientists in Whitehall (Manchester University Press, 1980).

Mercurial flights of Mariner 10

Joseph Veverka

Mercury: The Elusive Planet. By Robert G. Strom. *Smithsonian Institution Press/Cambridge University Press: 1987. Pp.197. \$19.95, £13.95.*

To most, Mercury is a drab, dead planet, superficially similar to the Moon. Unfortunately, this facile characterization of the Solar System's second smallest planet is as pervasive as it is inaccurate.

This small volume, the second in a new series on the Solar System published by the Smithsonian Institution Press, provides a good summary of current knowledge of a neglected member of the planetary family. The difficulties of observing Mercury from Earth are legendary, and it is not surprising that much of our detailed knowledge comes from the only spacecraft to have explored the planet, Mariner 10, which made three flybys of Mercury during 1974 and 1975.

This account, by a scientist intimately involved in all phases of the mission, can be divided into three parts: a review of pre-Mariner-10 ideas based on telescopic and radar observations, a summary of the Mariner 10 mission and its results, and finally a perspective of what Mercury is like and of the main problems that remain in deciphering the planet's current state and evolution.

Especially successful are those parts of the text which describe the author's own involvement: the account of the Mariner 10 mission and the discussion of the planet's geology. The presentation is well balanced, with only occasional lapses into

idiosyncrasy. It is not clear that understanding the origin of 'intercrater' plains is the *major* problem of Mercury's geology. Nor is it evident that crater statistics alone can reveal the detailed nature (size, composition and orbital characteristics) of bodies that have hit Mercury during its history.

Strom sets out to prove that Mercury is one of the more interesting places in the Solar System, and perhaps the key to understanding how the Earth-like planets evolved. The latter point is not developed convincingly, but the author never tires of underscoring Mercury's unique attributes (unexpectedly intense magnetic field, huge dense core, global system of thrust faults suggestive of crustal shrinking, and so on). Strom succeeds in making the case that the planet should not be neglected, but falls short of demonstrating his thesis that it is one of the most important and interesting bodies in the Solar System.

The text is essentially error-free, but the book has a few minor imperfections. Some figures are reproduced poorly and others are too small to be useful. At times, important concepts are explained inadequately: I doubt that a novice will understand why Mercury is in a 3:2 spin-orbit resonance rather than a 1:1 like our Moon. And I am certain that the inclusion of references would have helped the more serious reader.

Mercury: The Elusive Planet is successful as a popular account of the Mariner 10 mission and of current knowledge of Mercury. As such, anyone seeking a first exposure to these topics will find it well worth reading. □

Joseph Veverka is Professor of Astronomy and Space Sciences at Cornell University, Ithaca, New York 14853, USA.

The decipherment of Linear G

Sydney Brenner

Of Urfs and Orfs: A Primer on How to Analyze Derived Amino Acid Sequences. By Russell F. Doolittle. *University Science Books, 20 Edgehill Road, Mill Valley, California 94941/Oxford University Press: 1987. Pp. 103. Pbk \$14, £12.50.*

Nucleic Acid and Protein Sequence Analysis: A Practical Approach. Edited by M.J. Bishop and C.J. Rawlings. *IRL:1987. Pp. 417. Hbk £30, \$54; pbk £20, \$36.*

ABOUT eight months before those two amateurs, Watson and Crick, solved the structure of DNA in Cambridge, another amateur in the same place, Michael Ventris, announced that he had been able to decipher an ancient language, Linear B. Tablets written in this script had been unearthed in Knossos and had received many different interpretations. As it later turned out, all of these interpretations were quite fanciful; it seems that each of the scholars had fixed but different ideas about the language and simply bent the rules to suit prejudice (one even read the script from left to right or from right to left as suited his convenience). In this way lists of equipment in a quartermaster's store became transformed into dramatic announcements of the arrival of royal persons in awesome chariots. Ventris succeeded because he had the idea that Linear B was really Greek, something which his predecessors denied, and he was able to support this conclusion by a careful study of the internal evidence. The convincing coup was the easy and consistent translation of the Pylos tablets found in 1952, most of which were unromantic descriptions of business transactions in an outpost of the Mycenaean Empire.

I have always wondered whether there is more than just a joking resemblance between deciphering an ancient language and reading the genetic script. We can view Linear G, the language of DNA, as a very ancient script — the oldest there is, in fact — and some believe that there are clues to its meaning that derive from its origin and its evolution.

As we now know, DNA contains many different kinds of message written in the same alphabet of four letters, A, G, C and T, and the natural translation of each of these scripts requires a different machinery of molecular recognition. One language, defined by the triplet code, specifies the amino acid sequences of proteins, while others signal where genes begin and end and where their transcripts need to be spliced. There are the languages of gene regulation, of replication and of chromosome segregation. Indeed, a single

sequence could convey several messages, depending on how it is read. We can easily translate the coding sequences into the corresponding linear amino sequence, but the function of the protein is a property of its three-dimensional structure; we need to deduce this from the linear amino acid sequence and then read the surface of the protein to decipher the 'meaning' of a gene. This is our old friend, the still unsolved protein-folding problem. Of course, we can determine the three-dimensional structure of a protein empirically, but although methods are being continually improved, the task is still uncertain, arduous and expensive.

What encourages us now to approach the question is the torrent of sequence information resulting from technical advances in cloning and sequencing DNA molecules. Not so long ago, protein sequences could only be obtained by long and painful excavation, but now the databanks provide us with an archaeological treasurehouse of tablets of Linear G. A wide range of computational tools is now available to help us in sequencing and in interpreting the sequences, and the computer has increasingly become an indispensable tool in molecular biology laboratories. These two books are recently published guides to this important area.

Doolittle's little book *Of Urfs and Orfs* ("urfs" are unidentified reading frames, "orfs" are open reading frames) is a very readable primer and is strongly recommended to those entering the field. It explains how sequence homology can be established from sequence comparisons and how this may be used to understand the evolutionary connections between proteins. The section on how to make peptides and antibodies to them enables the student to materialize his protein out of the computer. The combination of computer algorithms, chemistry and immunology is a remarkable product of the technology revolution in biology.

Nucleic Acid and Protein Sequence Analysis is a practical "aid to biologists wishing to use computers for the acquisition, storage or analysis of nucleic acid and protein sequences". It is a reference book by many different authors, covering the whole range of the subject from computer hardware and available software to methods of sequence analysis and comparison. It has sections on protein structure prediction and on secondary structure analysis of nucleic acid sequences.

We are at the beginning of a new era in biology. The decipherment of Linear G — the generation of organisms from their genetic scripts — is the fundamental problem in biology. Learning how to do it is going to be a most challenging and exciting scientific venture. □

Sydney Brenner is Head of the Medical Research Council Molecular Genetics Unit, Hills Road, Cambridge CB2 2QH, UK.

Modest innovation

William Rhodes

Tributes to Paul Dirac. Edited by J.G. Taylor. Adam Hilger, Bristol/Taylor & Francis, Philadelphia: 1987. Pp.123. £9.95, \$22.

In April 1985 a memorial meeting for Paul Adrien Maurice Dirac was held at Cambridge University. *Tributes to Paul Dirac* is a collection of talks given at that meeting, together with reminiscences of Dirac by other colleagues and students. It is a eulogy to one of the greatest scientists the world has ever known.

The short period of time 1925–1928 experienced the birth of modern quantum theory. Three names stand foremost among the developers of this exciting new field: Heisenberg, Schrödinger and Dirac. All three formulated fundamental quantum mechanical equations. However, it was Dirac who quickly developed the unifying understanding of the structure of quantum theory and its connections with classical theory. His relativistic equation for the electron established the basis for quantum field theory, which now spearheads modern theoretical physics. In later years Dirac also made notable contributions to field theory, relativity and cosmology.

After a concise chronological sketch of the highlights of Dirac's personal and scientific life, the book is divided into two parts respectively containing personal reminiscences and accounts of his scientific contributions. The several historical photographs and the portrait of Dirac are excellent. The personal chapters, by students and colleagues, display varying degrees of closeness to Dirac and offer glimpses of a great man whose most outstanding personality traits were reticence and modesty. Although all of them clearly convey a mixture of affection and respect, only a few give insights into Dirac's nature. Particularly illuminating and valuable are the comments of A. Salam, J.E. Lannutti and S. Shanmugadhasan. In some cases, however, it seems as though the writers are searching for significance in small anecdotes.

Most of the chapters on Dirac's scientific contributions also contain personal recollections. The chapter by J. Mehra on Dirac's early work on quantum theory should be of interest to the non-specialist as well as the specialist. Dirac is said to have had little inclination towards matters of philosophy, but J.C. Polkinghorne's discussion of Dirac's interpretation of quantum mechanics brings out Dirac's deep interest in the nature of physical reality. Dirac's persistent view that "physical laws should have mathematical beauty" (a statement he wrote on the

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Paul Dirac — "one of the greatest scientists the world has ever known".

blackboard at Moscow University in 1955) certainly has ontological content. But a few of the scientific presentations, all on subjects to which Dirac made notable contributions, are quite technical and are understandable only to the specialist.

In reading this book, one must realize that it is a tribute, not a definitive book on the life, character and scientific works of Paul Dirac. The world awaits a detailed and objective study of this outstanding figure of scientific history. Meanwhile, *Tributes* will serve as a pleasant introduction to those who were not fortunate enough to have known him. □

William Rhodes is a Professor in the Department of Chemistry, Florida State University, Tallahassee, Florida 32306, USA.

• Cambridge University Press has just published a further collection of essays as a memorial to Dirac. Entitled *Paul Adrien Maurice Dirac: Reminiscences About a Great Physicist*, the book begins frustratingly, with an assortment of personal recollections that are occasionally interesting enough as biography but which, when anecdotal, tend to fall flat. Something in Dirac's persona, with his famous brevity of utterance, may have charged these incidents with a particular significance but it barely survives the retelling.

In contrast, the inspirational influence of his science and of the elegant way in which he communicated it are compellingly recounted by a variety of eminent ex-students and colleagues. Most interestingly, there is a posthumous contribution by Dirac himself setting out in plain and uncompromising language his fundamental objections to the very successful expedient of renormalization, together with a more technical but enjoyably informal survey of supersymmetry and superstring theories which, as Abdus Salam describes, better satisfy the criterion of mathematical beauty on which Dirac insisted. Price of the book is £30, \$49.50.

Philip Campbell

Genetic reference

Dorothy A. Miller

Human Genetics: Problems and Approaches, 2nd Edn. By F. Vogel and A. G. Motulsky. Springer-Verlag:1987. Pp.807. DM 148.

Human Genetics was first published in 1979, and was a vast and no doubt daunting undertaking for Vogel and Motulsky. The success of the book, coupled with the many and often rapid changes that have taken place in the subject, have now induced them to carry out the no less daunting task of creating a revised edition. Those who liked the first edition will be pleased with the additions. Those who are not yet familiar with the volume have a treat in store.

The authors retain a scholarly approach. They cite more than 2,400 references, a sizeable proportion of them from 1980 or later. There are multiple tables covering such diverse subjects as the proportion of patients affected by new mutations in autosomal dominant diseases, the average stature of Scandinavian adult males from the Stone Age through to 1939, and the genetic diseases for which carrier detection is advisable. The text is particularly strong in dealing with the quantitative aspects of human genetics, and is studded with cautions regarding interpretation of data and suggestions for future research.

As well as giving the facts, the authors have not avoided commenting on controversial issues. For example, they present a stern view of the role of human geneticists in abetting the excesses of Nazi Germany, and they argue that therapy should not be used with fertilized human eggs. On the other hand they are optimistic regarding the prospects for somatic gene therapy.

As must be expected in a volume that covers so large an area, there are errors both of commission and omission. Some are minor, such as the interchange of R-banded chromosomes 4 and 5 in Fig. 2.9. Others are more serious because they mislead the reader. The typographical error which means that EpG (rather than CpG) is cited as the methylated dinucleotide is unfortunate, because this is the only reference to this important sequence. The elegant figure in which Cavanee *et al.* presented mechanisms for deriving homozygosity of a proposed retinoblastoma gene has been redrawn in such a way that now neither gene conversion nor mitotic recombination is illustrated. The subject of dermatoglyphics is virtually ignored. One exception is that a list of the main clinical features of trisomy 18 includes arches on "three or more finger tips"; the much more striking fact is that 80 per cent of these individuals have arches on seven or more finger tips,

whereas most normal individuals have no more than one. More generally, the index is inadequate. This is important because the treatment of subjects is fragmented and readers would be well advised to add their own annotation.

Molecular biology is not as well presented as one would like, considering the impact this discipline is having on genetics. The figure designed to illustrate the use of restriction fragment length polymorphisms (RFLPs) to distinguish between homozygote and heterozygote shows flow-sorted chromosomes as the starting material, whereas the strength of the method lies in the ability to use small amounts of genomic DNA from almost any source. *MspI* and *TaqI* are cited as

particularly useful enzymes for finding DNA variants, but there is no indication that this is so because the methylation of CpG in their recognition sequences can lead to mutation of C to T.

In my view *Human Genetics* is not so much a textbook as a wonderful reference source for teachers, researchers, genetic counsellors and students. As a bonus it is also a contemporary commentary on the state of the field by two expert practitioners. The bottom line is that, despite some imperfections, you will of course want a copy. □

Dorothy A. Miller is a Professor in the Department of Molecular Biology and Genetics, Wayne State University, Detroit, Michigan 48201, USA.

On the road

Joseph Palca

Flattened Fauna: A Field Guide to Common Animals of Roads, Streets, and Highways. By Roger M. Knutson. Ten Speed Press, PO Box 7123, Berkeley, California 94707:1987. Pp.88. Pbk \$4.95.

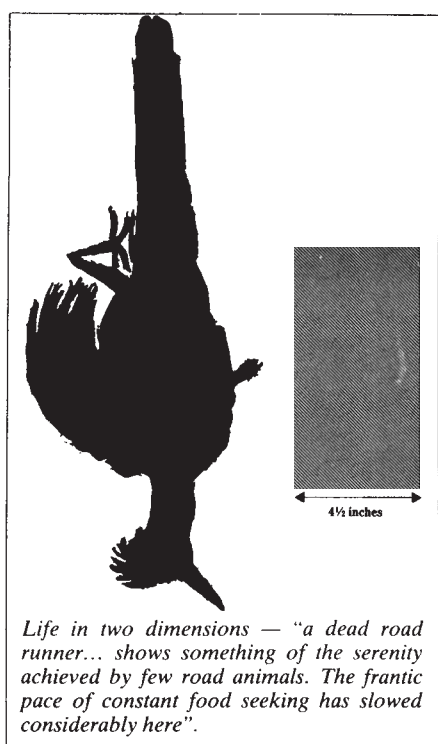
THE American love affair with the automobile has, for many, obscured an appreciation of the richness of fauna to be found in the United States. Although the interstate highway system has opened new vistas to the ever-growing number of urban dwellers, driving in a steel and glass capsule with air conditioner and stereo turned up high makes observing nature difficult.

But our notion of savouring the outdoors has been strongly determined by naturalists such as James Audubon and John Muir, whose views were shaped by the technology of a different era. A new approach could well help unite what the planet has to offer with what civilization has led us to expect.

In this regard, Roger Knutson has written what may be a seminal work. Without apologies, he offers a valuable reference work for both the casual vacationer and the serious student for making sense of the intersection between car and animal. Armed with *Flattened Fauna*, "a Sunday drive can become a safari into a new habitat populated with animals unlike those you have seen before".

There is a dichotomous key to help novices find their way through the complexities of observing animals that have died on the road. Silhouettes are provided throughout the book as an aid to identification, though in the case of snakes these are not much help.

Observing road fauna requires its own rules, and Knutson is not altogether successful when he tries to build on the principles of classical field zoology. First



Life in two dimensions — "a dead road runner... shows something of the serenity achieved by few road animals. The frantic pace of constant food seeking has slowed considerably here".

explaining the importance of mimicry for species survival, he then states "whether the flattened sod clump mimics the porcupine or vice versa is a moot point for the student of road animals, interesting though it may be for the evolutionary biologist". But he also tantalizes the reader by noting that "a rusted hubcap looks enough like a painted turtle to confuse even a longtime student of road artifacts". Can this really be a coincidence, or are there more subtle factors at work here?

Knutson gives due credit to James Simmons's *Feathers and Fur on the Turnpike* (1938) as a pioneering work in the study of road fauna. Fifty years on we must hope that his own book attracts renewed interest and funding to this fascinating but neglected area of the biological sciences. □

Joseph Palca is Washington News Editor of Nature.

Cancer near nuclear installations

David Forman, Paula Cook-Mozaffari, Sarah Darby, Gwyneth Davey, Irene Stratton, Richard Doll & Malcolm Pike

There has been no general increase in cancer mortality near nuclear installations in England and Wales during the period 1959–1980. Leukaemia in young people may be an exception, though the reason remains unclear.

THE Office of Population Censuses and Surveys (OPCS) has recently published a report "Cancer Incidence and Mortality in the Vicinity of Nuclear Installations England and Wales, 1959–80"¹. The report is very detailed and difficult to assimilate. We have, therefore, summarized here what we believe to be the principal results and the main conclusions that can be drawn from them.

The OPCS report

The installations included in the report are shown in Table 1. All pre-1974 Local Authority Areas (LAAs) with at least one-third of their population living within 10 miles of any of these installations were included in the analysis. The report provided standardized mortality and registration ratios (SMRs and SRRs) for 25 specific and combined cancer sites for groups of these LAAs around each installation using appropriate regional rates (age, sex and year specific rates for the Standard Region², excluding conurbations, in which the installation was situated) for standardization. These SMRs and SRRs were presented for both sexes combined and essentially for three discrete age groups: 0–24, 25–74 and 75+ years, although the information was also made available for the age group 0–9 years. The report also compared the trends in SMRs and SRRs both over time and with increasing proximity to the installations.

In addition, comparisons were made between the SMRs and SRRs of the grouped LAAs around each installation and the SMRs and SRRs of groups of control LAAs. Each group of control LAAs was constructed by matching each installation LAA with a non-installation LAA, selected, where possible, to be in the same Standard Region² and have a similar urban–rural status and population size as its comparison LAA. For LAAs that were County Boroughs, control selection was also based on comparable social class structure when there was a choice within the above constraints. The rates for the same Standard Region were used for calculating the SMRs and SRRs for each installation LAA and for its control.

In order to focus attention on overall patterns the report provided summed results for the LAAs in two mutually

exclusive combined groups of installations and the respective control LAAs. These were, firstly, all the installations that started operation before 1955, except Sellafield (the "pre-1955 installations"), and secondly, all the CEBG installations combined, see Table 1.

Four distance zones were considered in the report: (1) LAAs with at least two-thirds of their population within 6 miles of an installation; (2) LAAs with at least two-thirds of their population within 8 miles of an installation; (3) LAAs with at least two-thirds of their population within 10 miles of an installation; and (4) LAAs with at least one-third of their population within 10 miles of an installation.

These cumulative zones are referred to as cumulative zones 1, 2, 3 and 4. For some analyses mutually exclusive ("discrete") distance zones were used; these discrete zones were constructed from the cumulative zones by successively excluding inner zones, for example discrete zone 2 = zone 2 – zone 1, and so on. It should be noted that LAAs in the vicinity of all installations are not represented in every distance zone. In particular, the innermost distance zone, as defined above, contains no LAAs that are near to Sellafield, Aldermaston, Berkeley/Oldbury and Hinkley.

The report provided data for the 22 years 1959–1980, a period that commenced several years after the start of operations at Sellafield and at each of the group of pre-1955 installations. The CEBG installations, however, began

operations between 1961 and 1971, and Winfrith became operational in 1964. Data for these installations were also available in the report from the year subsequent to start-up for some types of cancer.

In addition to the analyses of LAAs around installations, the report also presented results for a group of coastal LAAs to the north and south of Sellafield to investigate whether seaborne discharges from Sellafield might be affecting cancer rates. These coastal LAAs have some overlap with the LAAs around Sellafield, Springfields, Capenhurst and Wylfa. Control LAAs were selected for these coastal LAAs using the criteria stated above.

Methods

The principal results, in our opinion, are those relating to mortality at ages 0–24 and 25–74 years for combined groupings of installations, categorized by likely environmental impact or similar start-up date. Other data are less relevant, because:

(a) Cancer registration data are of variable quality³ and the report itself highlighted a number of inconsistencies in the registration information^{1,4}. In particular, the ratio of the number of cancer registrations to the number of cancer deaths was higher in the LAAs around pre-1955 installations than in the corresponding control LAAs, which suggests a general registration bias in favour of the installation areas.

(b) The mortality results for the age group 75+ are less reliable than those for

Table 1 Nuclear installations and summary groupings

Installation	Summary grouping	Date of first criticality or waste release*
BNFL Sellafield	Pre-1955 (excl. Sellafield)	1950
BNFL Springfields		1948
BNFL Capenhurst		1953
Amersham International plc		1946
MOD Aldermaston		1952
UKAEA Harwell		1947
UKAEA Winfrith	CEBG	1964
CEGB Bradwell		1961
Berkeley & Oldbury		1961
Hinkley		1964
Trawsfynydd		1964
Dungeness		1965
Sizewell		1965
Wylfa		1971

* Referred to as "start-up date" in the text.

Table 2 Comparison of standardized mortality ratios (SMR) between groupings of installation LAAs and their controls*

Category of cancer	Installation grouping	Cumulative distance zone†		Installation LAAs		Control LAAs		Relative risk‡	P
		(a)	(b)	Obs	SMR	Obs	SMR		
a Persons aged 0-24									
Lymphoid leukaemia	Pre-1955 (excl. Sellafield)	1	2	42	113.4	20	54.1	2.10	0.004
	All installations	1	—	44	112.1	22	56.1	2.00	0.005
All brain	Sellafield	2	—	5	144.6	0	0.0	∞	0.033
b Persons aged 25-74									
Liver	All CEEGB	2	3, 4	29	189.0	10	62.4	3.03	0.001
	Winfrith	4	—	16	127.0	6	52.0	2.44	0.043
	All installations	4	2, 3	395	100.0	313	81.3	1.23	0.004
Lung	All CEEGB	4	3	2,956	95.0	3,134	90.3	1.05	0.025
Hodgkin's disease	Pre-1955 (excl. Sellafield)	1	2, 3, 4	141	95.9	104	70.7	1.36	0.011
	All installations	1	2, 3	149	96.7	112	72.4	1.34	0.010
All lymphomas	Pre-1955 (excl. Sellafield)	1	—	382	98.5	338	86.2	1.14	0.040
Unspec. brain and CNS	Pre-1955 (excl. Sellafield)	4	—	357	109.2	302	95.7	1.14	0.049
All malignancies	All CEEGB	4	3	10,855	98.4	11,536	94.5	1.04	0.001
	Sellafield	4	2, 3	2,399	97.7	2,491	89.8	1.09	0.002

* All comparisons where the SMR for the installation LAAs was significantly greater than for the control LAAs ($P < 0.05$, one-sided test) in any of the four cumulative distance zones are indicated. Detailed results are given for the distance zone with the most extreme significance level. Sellafield does not appear on this table for leukaemia in the 0-24 age-group because the increase that has been reported was confined to Millom Rural District. In this analysis Millom Rural District is in distance zone 4 where it is combined with Ennerdale Rural District (also in distance zones 2 and 3) and Whitehaven Metropolitan Borough (also in distance zone 3). No increase in leukaemia mortality was found in these other areas.

† (a) With most extreme significant difference (lowest probability); (b) with significant differences ($P < 0.05$).

‡ Relative risk = SMR (installation LAAs)/SMR (control LAAs).

other ages because diagnostic accuracy on death certificates diminishes with age^{1,5}.

(c) The results for LAAs around individual installations are based on small numbers and are likely to be less informative than those for groups of similar installations. We have, therefore, mainly used the results for combined installation groupings provided in the report. We have analysed data for grouped LAAs around:

(i) The pre-1955 installations excluding Sellafield. Although the activities carried out in these installations are heterogeneous, the grouping combines all the establishments other than Sellafield which, by 1980, had been operating for more than 25 years and around which any long-term effects could most easily be identified.

(ii) Combined CEEGB installations. All are thought to have a similar impact on their local environment, in terms of their radioactive discharges⁶.

(iii) Sellafield. Sellafield first released waste in 1950, but a specific problem had already been identified in relation to it before the OPCS study⁷, and it also needs to be considered separately because the radiobiological impact of the discharges from Sellafield is likely to have been far in excess of that from any other establishment^{7,8}.

(iv) Winfrith. Winfrith first released waste in 1964, and does not fit into any of the other categories.

(v) All installations grouped together, obtained by summing (i) to (iv) above.

We also analysed results for the group of coastal LAAs to the north and south of that part of mid-Cumbria in which Sellafield is situated. The three LAAs in mid-Cumbria have been excluded as they are already included in the Sellafield group (group (iii) above) and Liverpool County Borough was excluded as it has a very

different population density and social class composition from the remaining coastal LAAs. The final group consisted of 58 coastal LAAs.

The 22 "types" of cancer included in the analysis are listed in the Appendix (p. 505). The SMRs given in the report are comparisons with regional standards which make allowance for gross regional and urban-rural variation in cancer rates. However, any remaining difference in the demographic or socio-economic structure of the LAAs around an installation compared to that of a region as a whole will result in an effect on the SMR unrelated to the presence of the installation. This type of effect could be expected to be minimized in the comparison between grouped installation LAAs and grouped control LAAs, so we have largely restricted attention to such comparisons.

Our approach, therefore, has been to compare mortality rates in the above installation groupings within each of the four distance zones and in the coastal LAAs around Sellafield with the rates in their respective control LAAs. Data for the CEEGB installations and Winfrith were considered only subsequent to start-up. (In fact, because the OPCS report did not present all the necessary information, it was not possible to divide the data into pre- and post-start-up for certain cancer sites. In these instances data were used from 1966-1980 instead, see Appendix.) The numbers involved in these instances are small and will not affect the overall pattern of occurrence.

For all the 22 types of cancer, the SMRs are presented in (or can be calculated from) the OPCS report for each grouping of installation LAAs that we have considered and for their control LAAs. We have calculated the ratios of the SMRs for the grouped installation LAAs to the

SMRs for the grouped control LAAs and called these ratios "relative risks" (RRs). For each cumulative distance zone, one-sided significance tests have been performed to test the hypothesis that the RR for the grouped installation LAAs is greater (or less) than unity. These tests were performed on the observed numbers of deaths which we have assumed follow Poisson distributions with means proportional to the number of deaths expected. Tests for trend in the RRs in the discrete distance zones with increasing proximity to an installation were also carried out using a likelihood ratio test.

Results

All comparisons which lead to the RR for the grouped installation LAAs being significantly greater than unity (one-sided $P < 0.05$) in any of the four cumulative distance zones are shown in Table 2 ("positive" results). Table 3 corresponds to Table 2, and shows all comparisons which lead to the RR for the grouped installation LAAs being significantly less than unity ("negative" results). All significant differences between the grouped coastal LAAs and their control areas, irrespective of whether the RRs are greater or less than unity are shown in Table 4.

In Tables 2 and 3 we show all the distance zones in which significant differences are observed between the SMRs in the installation and control areas. We also show the observed number of deaths and the SMRs for the installation and control LAAs in the distance zone which shows the most extreme difference in cancer mortality rates in terms of its statistical significance.

Numerically there are substantially more negative results listed in Tables 3 and 4b than there are positive results listed in Tables 2 and 4a and a greater variety

Table 3 Comparison of standardized mortality ratios (SMR) between groupings of installation LAAs and their controls*

Category of cancer	Installation grouping	Cumulative distance zone†		Installation LAAs		Control LAAs		Relative risk‡	P
		(a)	(b)	Obs	SMR	Obs	SMR		
a Persons aged 0-24									
Myeloid leukaemia	Winfrith	4	—	2	55.0	9	234.9	0.23	0.042
Malignant brain	All CEGB	2	—	5	69.6	15	204.0	0.34	0.025
All brain	All CEGB	2	—	6	64.5	16	167.3	0.39	0.032
Malignant brain	All installations	1	—	34	80.2	51	118.1	0.68	0.050
Non-Hodgkin's lymphoma	All CEGB	2	—	0	0.0	5	147.9	0.00	0.034
All lymphomas and leukaemia	Winfrith	4	—	16	94.3	30	168.7	0.56	0.039
All malignancies	All CEGB	2	—	49	97.7	72	138.6	0.70	0.035
b Persons aged 25-74									
Myeloid leukaemia	Pre-1955 (excl. Sellafield)	3	4	473	97.7	522	114.2	0.86	0.008
	All installations	3	4	575	96.7	651	113.2	0.85	0.003
Non-lymphoid leukaemia	Pre-1955 (excl. Sellafield)	3	4	576	97.5	621	111.4	0.87	0.011
	All installations	3	4	701	96.8	772	110.3	0.88	0.007
Lymphoid leukaemia	Winfrith	3	4	2	33.4	12	203.8	0.16	0.007
All leukaemia	Winfrith	3	—	21	87.5	36	150.0	0.58	0.032
	All CEGB	1	2, 3	10	50.1	28	118.6	0.42	0.012
	All installations	1	—	379	89.8	437	101.8	0.88	0.040
Malignant brain	Pre-1955 (excl. Sellafield)	3	4	1,116	95.0	1,203	107.5	0.88	0.002
	All CEGB	1	—	12	67.6	28	132.8	0.51	0.033
	All installations	3	4	1,333	94.3	1,474	107.2	0.88	<0.001
Benign brain	Pre-1955 (excl. Sellafield)	4	3	152	88.6	185	112.5	0.79	0.016
	All CEGB	1	—	0	0.0	7	324.4	0.00	0.014
	All installations	4	2, 3	183	88.0	221	109.0	0.81	0.018
All brain	Pre-1955 (excl. Sellafield)	3	4	1,523	96.7	1,605	107.0	0.90	0.002
	All CEGB	1	—	15	65.0	38	138.7	0.47	0.008
	All installations	3	4	1,811	96.1	1,952	106.6	0.90	<0.001
Lung	Pre-1955 (excl. Sellafield)	2	3, 4	7,500	96.0	7,749	103.8	0.93	<0.001
	Winfrith	3	—	288	101.4	330	119.1	0.85	0.025
	All installations	2	3, 4	8,765	96.3	9,059	102.6	0.94	<0.001
Bone	Sellafield	3	2, 4	4	49.1	16	182.0	0.27	0.011
	All CEGB	4	3	32	88.7	54	136.8	0.65	0.032
	All installations	3	4	231	89.7	277	112.8	0.80	0.006
Multiple myeloma	Pre-1955 (excl. Sellafield)	1	—	146	90.3	187	113.6	0.80	0.021
	Winfrith	4	—	25	75.2	39	130.0	0.58	0.021
	All installations	1	2	153	88.5	199	112.1	0.79	0.016
Testis	All CEGB	4	—	18	62.3	37	127.0	0.49	0.008
All lymphomas and leukaemia	All CEGB	1	2, 3	26	65.0	55	116.4	0.56	0.009
All malignancies	Pre-1955 (excl. Sellafield)	2	1, 3, 4	25,928	96.6	26,074	101.4	0.95	<0.001
	Winfrith	3	—	1,055	102.7	1,142	113.3	0.91	0.011
	All installations	2	1, 3, 4	30,645	97.5	30,835	101.1	0.96	<0.001

* All comparisons where the SMR for the installation LAAs was significantly less than for the control LAAs ($P < 0.05$, one-sided test) in any of the four cumulative distance zones are indicated. Detailed results are given for the distance zone with the most extreme significance level.

† (a) With most extreme significant difference (lowest probability); (b) with significant differences ($P < 0.05$).

‡ Relative risk = SMR (installation LAAs)/SMR (control LAAs).

of different cancer types are included in the negative results. In this article we are not, however, concerned with the reasons why the risk of death from a number of cancer types is relatively low in the installation areas, except in so far as they help to interpret the opposite findings. We have, therefore, examined further only those types of cancer that had a relatively higher mortality in LAAs around the installations than in the control LAAs.

Cancer at ages 0-24. Two types of cancer have caused death relatively more frequently in persons under 25 years of age around some of the installations than in the corresponding control areas: namely, lymphoid leukaemia and all brain tumours. The RR for lymphoid leukaemia is significantly elevated in cumulative zones 1 and 2 in the LAAs grouped around the pre-1955 installations. As the populations in the LAAs around the pre-1955 installations constitute 81 per cent of the entire study population, this results in a significant elevation of lymphoid leukaemia, though in zone 1 only, around

all installations combined. These findings are shown in Table 5 in the context of the overall results for leukaemia (lymphoid, non-lymphoid and all types) in the four discrete distance zones. For lymphoid leukaemia there is a significantly increasing trend in RR with increasing proximity to an installation in the pre-1955 group (two-sided $P = 0.048$) and a significantly decreasing trend with increasing proximity to an installation for the grouped

CEBG installations (two-sided $P = 0.015$).

The RR for all brain cancer (that is, for brain tumours described as malignant, benign or of unspecified malignancy) is significantly elevated in cumulative zone 2 around Sellafield. When data for this type of cancer are examined in the four discrete zones around all the installation groupings, two comparisons show statistically significant trends. One is of an increasing risk with increasing proximity

Table 4 Comparison of standardized mortality ratios (SMR) between the coastal grouping of LAAs† and the control grouping*

Persons aged 25–74 Category of cancer	Coastal LAAs Obs	LAAs SMR	Control LAAs Obs	LAAs SMR	Relative risk‡	<i>P</i>
<i>a</i> Coastal LAAs greater than control LAAs						
Multiple myeloma	659	107.0	518	96.6	1.11	0.043
<i>b</i> Coastal LAAs less than control LAAs						
Bone	249	89.6	284	114.1	0.79	0.003
All malignancies	68,433	97.1	60,935	98.4	0.99	0.009

* All comparisons where the SMR for the coastal grouping was significantly greater or less than the controls ($P < 0.05$, one-sided test) are included.

† Excluding mid-Cumbria LAAs and Liverpool County Borough.

‡ Relative risk = SMR (coastal LAAs)/SMR (control LAAs).

Table 5 Standardized mortality ratios (SMR) for leukaemia in installation LAAs and relative risks (RR) of installation LAAs to control LAAs, 0-24 years age group, discrete distance zones

		Discrete distance zone				Sign [§]	Trend Probability
Grouping		1	2	3	4		
a Lymphoid leukaemia (1968-1980)¶							
Pre-1955	SMR	113.4	109.1	103.8	91.29		
	RR	2.10*	0.89	0.98	0.94	+	0.048
All CEBG	SMR	102.1	76.8	106.4	162.0		
	RR	1.05	0.34	1.54	2.96†	—	0.015
Winfrith	SMR	0.0	— [‡]	44.8	123.7		
	RR	— [#]	—	0.29	1.32	—	0.297
Sellafield	SMR	— [‡]	138.5	53.8	333.7		
	RR	—	1.11	0.28	∞**	—	0.276
All installations	SMR	112.1	105.5	101.0	113.2		
	RR	2.00*	0.75	0.96	1.30	+	0.425
b Non-lymphoid leukaemia (1968-1980)¶							
Pre-1955	SMR	110.0	119.8	86.1	125.8		
	RR	0.97	1.10	0.71	1.40	—	0.786
All CEBG	SMR	0.0	118.0	75.6	151.1		
	RR	0.00	1.00	5.85	1.65	—	0.121
Winfrith	SMR	0.0	— [‡]	110.1	75.4		
	RR	— [#]	—	1.17	0.28	+	0.340
Sellafield	SMR	— [‡]	61.0	70.2	565.0*		
	RR	—	1.09	∞**	4.00	—	0.412
All installations	SMR	107.0	114.9	84.8	137.4		
	RR	0.91	1.08	0.84	1.28	—	0.478
c All leukaemia (1959-1980)¶							
Pre-1955	SMR	105.9	109.4	93.5	95.5		
	RR	1.20	0.98	0.80	0.94	+	0.094
All CEBG	SMR	130.5	98.6	104.0	131.7		
	RR	1.10	0.55	1.50	1.59	—	0.054
Winfrith	SMR	0.0	— [‡]	104.7	113.6		
	RR	— [#]	—	0.82	0.56	+	0.696
Sellafield	SMR	— [‡]	102.5	83.4	263.2†		
	RR	—	0.68	0.47	3.94	—	0.076
All installations	SMR	106.3	107.1	94.8	106.9		
	RR	1.19	0.86	0.84	1.02	+	0.382

* $P < 0.01$ (one-sided test for increase above 100 (SMR) or unity (RR)).† $P < 0.05$ (one-sided test for increase above 100 (SMR) or unity (RR)).

‡ No LAAs in this distance zone.

§ Sign of trend in RR with increasing proximity to installation.

|| Probability of trend arising due to chance (two-sided test).

¶ Data for lymphoid and non-lymphoid leukaemia are only separately available since the introduction of the eighth revision of the International Classification of Diseases in 1968. Data for all leukaemia are presented for the entire time period 1959-1980, but are only included since start-up of the CEBG installations and since 1966 for Winfrith.

No deaths in installation or control LAAs. ** No deaths in control LAAs.

to Sellafield (two-sided $P = 0.026$). The other is of an increasing risk with decreasing proximity to the CEBG installations (two sided $P = 0.013$).

Cancer at ages 25-74. Seven types of cancer have caused death relatively more frequently in persons aged 25-74 years around some of the installation groupings or in the grouped coastal LAAs: namely, Hodgkin's disease, all lymphomas, multiple myeloma, liver cancer, lung cancer, tumours of the brain and central nervous system of unspecified malignancy, and all malignancies.

The mortality from Hodgkin's disease is elevated in all four cumulative zones in the LAAs around the pre-1955 installations in comparison with their control LAAs. This leads to elevated RRs in cumulative zones 1, 2 and 3 for all installations combined. The results for the discrete distance zones are given in Table 6. Statistically significant trends are seen for an increasing relative risk with increasing proximity to an installation for all installations combined (two-sided $P = 0.040$) and for Winfrith (two-sided $P = 0.002$).

Hodgkin's disease is a major component in the category of all lymphomas, the mortality from which is also elevated in distance zone 1 around the pre-1955 installations.

The mortality from multiple myeloma is significantly elevated in the coastal LAAs, but not in any of the LAAs around the nuclear installations. Because of the finding of an increased risk of myeloma with increasing dose of radiation in the Sellafield⁹ and Hanford^{10,11} workers, the results are examined in detail for the LAAs in each of the discrete distance zones around the grouped installation LAAs in Table 7. There is a statistically significant decrease in relative risk with increasing proximity to an installation (two-sided $P = 0.017$) for the LAAs around the pre-1955 installations.

Liver cancer RRs are elevated in cumulative zones 2, 3 and 4 around the grouped CEBG installations and in cumulative zone 4 around Winfrith. These results are reflected in elevated RRs for this cancer in cumulative zones 2, 3 and 4 around all installations combined.

The other positive results around the CEBG installations are those for lung cancer and for all malignancies in cumulative zones 3 and 4. The RR for all malignancies is also elevated in all three distance zones around Sellafield. However, both of these cancer categories have significantly lower RRs around the grouped pre-1955 installations and around all installations combined (Table 3).

The results for liver cancer, lung cancer and all malignancies for all four cumulative zones around the CEBG installations are given in Table 8, which also shows separate data for each zone before and after start-up of the installations. For liver cancer the excesses in zones 2, 3 and 4 were all present before the opening of the installations as well as after, and the earlier SMRs were, in each case, higher than subsequently. For lung cancer and for all malignancies, the RRs were all close to unity both before and after start-up, the highest being 1.06. In cumulative zones 3 and 4, the RRs and the SMRs tended to be higher in the installation LAAs than in the control LAAs after the opening of the installations than before, but the reverse was true for cumulative zones 1 and 2.

The remaining positive result is the increased RR for tumours of the brain and central nervous system of unspecified malignancy in cumulative zone 4 around the pre-1955 installations and around all installations combined. This was more than compensated for by significantly decreased RRs for the other three brain cancer categories (malignant brain cancer, benign brain cancer and all brain cancer) in these installation groupings (Table 3).

Discussion

The comparison of Tables 2 and 4a with Tables 3 and 4b shows that the SMRs for LAAs in the vicinity of nuclear installations or in the 58 selected coastal LAAs are significantly less than the SMRs in their control LAAs more often than the reverse. This, moreover, remains true if attention is concentrated on those types of cancer that have been particularly associated with exposure to ionizing radiation¹²; namely, leukaemia, bone cancer and multiple myeloma. This provides strong evidence that there is no generalized increase in cancer mortality around nuclear installations in England and Wales either in young persons or in adults.

Although many of these results will be chance findings as so many comparisons were made, the tendency for cancer mortality to be lower near nuclear installations than in the control LAAs is too strong to be explained solely on these grounds. We do not think that radioactive discharges can protect against the development of cancer and we conclude that there are likely to have been sufficiently large socio-economic or environmental differences between the installation and the control

Table 6 Standardized mortality ratios (SMR) for Hodgkin's disease in installation LAAs and relative risks (RR) of installation LAAs to control LAAs, 25-74 years age group, discrete distance zones

Grouping		Discrete distance zone				Sign [§]	Trend Probability
		1	2	3	4		
Pre-1955	SMR	95.9	97.9	99.7	103.5	+	0.102
	RR	1.36†	0.99	1.12	0.96		
All CEBG	SMR	63.5	127.4	112.0	79.4	+	0.966
	RR	0.64	0.96	1.34	0.78		
Winfrith	SMR	540.5*	—‡	93.3	101.6	+	0.002
	RR	∞**	—	1.24	0.56		
Sellafield	SMR	—‡	77.1	64.3	92.8	+	0.643
	RR	—	1.05	0.77	0.74		
All installations	SMR	96.7	100.5	99.7	99.2	+	0.040
	RR	1.34†	0.98	1.14	0.89		

* $P < 0.01$ (one-sided test for increase above 100 (SMR) or unity (RR)).† $P < 0.05$ (one-sided test for increase above 100 (SMR) or unity (RR)).

‡ No LAAs in this distance zone.

§ Sign of trend in RR with increasing proximity to installation.

|| Probability of trend arising due to chance (two-sided test). ** No deaths in control LAAs.

LAAs to have had an effect on the pattern of cancer mortality, despite the efforts that were made to choose pairs of LAAs that were socially similar. It follows that the excesses of mortality from some types of cancer in the installation LAAs, if they are not due to chance, may also result from these differences. It is, therefore, necessary to consider carefully whether each of the positive results may be due to chance, or to socio-economic/environmental differences, or to the direct presence of the installations.

We have concentrated initially on the age group 0-24 years, as it might be expected that any effect of the installations on cancer mortality in the general population would first be seen in young people. Studies of the effects of high doses of radiation have shown that children have a greater proportionate increase in risk following a given level of exposure than adults¹³. In addition, for inhaled and ingested radionuclides, children are thought to receive higher doses than adults from similar levels of environmental contamination, under some circumstances¹⁴.

In the age group 0-24 years there are two results of possible concern, those for lymphoid leukaemia and for all brain cancer. In the period 1968-1980, the RRs for lymphoid leukaemia are raised in the LAAs around the pre-1955 installations in cumulative zones 1 and 2, and this is reflected

in a raised RR around all installations combined in cumulative zone 1. The differences observed between the SMRs for lymphoid leukaemia in the installation LAAs and the control LAAs are the most highly significant of all the differences in this age group.

Several factors argue against the observed excess of lymphoid leukaemia being a chance association. Most of the evidence for the excess comes from the LAAs around the pre-1955 installations, and any increase in risk that was actually caused by the installations would have had the greatest time to accumulate in these LAAs. For the pre-1955 installations, the increased RR is restricted to those LAAs closest to the installations, that is, in zone 1 (Table 5); and each of the four individual pre-1955 installations with LAAs in zone 1 (Springfields, Capenhurst, Amersham and Harwell), shows an elevated RR in this zone, although for Capenhurst and Harwell the individual increases do not reach statistical significance (Table 9). There is also a suggestion of an increase of lymphoid leukaemia in the coastal areas (RR 1.24, one-sided $P = 0.115$). Because the Sellafield discharges are far in excess of those from any other establishment and a substantial proportion of those discharges are made into the sea, these coastal areas could be expected to be the first to experience any widespread effects

associated with the discharges from the plant.

Moreover, because of particular interest in the effects of radiation on very young children¹², we also looked at the lymphoid leukaemia results for the age group 0-9 years, which are available in the OPCS report on microfiche. In this age group all of the above effects were stronger. For the LAAs in zone 1 around the pre-1955 installations the RR for lymphoid leukaemia is 3.95 (one-sided $P = 0.001$) in comparison with 2.10 (one-sided $P = 0.004$) for the age group 0-24 years. There is also a significant positive trend with increasing proximity to the installation (two-sided $P = 0.035$). Finally in the coastal LAAs there is a significantly elevated RR of 1.42 (one-sided $P = 0.038$).

It should be noted that the excesses of lymphoid leukaemia in the LAAs around the pre-1955 installations depend, in large part, on particularly low SMRs in the control LAAs in zone 1. Thus the combined SMR for the age group 0-24 years in the LAAs in zone 1 around the pre-1955 installations is 113.4, whereas that for the corresponding control LAAs is 54.1. For the age group 0-9 the corresponding SMRs are 113.7 and 28.8. The reason for these low control SMRs is unclear, and demonstration of equivalent low SMRs in other possible control LAAs is needed before it can be concluded confidently that the resulting elevated RRs have aetiological significance.

The difference between the results for lymphoid and non-lymphoid leukaemia, seen in Table 5, means that the combined category of all leukaemia shows no increased risk in the coastal LAAs and only a slight indication of an increased RR in zone 1 around the pre-1955 installations (RR = 1.20, one-sided $P = 0.116$). However, the results for lymphoid leukaemia are based on mortality since 1968 (data on leukaemia subtypes were not available before then). The RR for all leukaemia in zone 1 since 1968 is 1.46 and is statistically significant (one-sided $P = 0.029$). Therefore, by restricting attention to the more reliable all-leukaemia category and to the later time period, 1968-1980, there is still reasonable evidence of an elevated risk in the age group 0-24 years. In the younger age group, 0-9 years, the RR is somewhat higher at 1.66, and is of borderline significance ($P = 0.051$). In the coastal LAAs the RRs for all leukaemia in this later period are 1.10 ($P = 0.181$) at ages 0-24 and 1.26 ($P = 0.072$) at ages 0-9. Whether or not these increases are particularly associated with lymphoid leukaemia, as our initial results suggest, is open to question and depends on explaining why the SMRs in the control LAAs are especially low for this type of cancer and not for non-lymphoid leukaemia.

The present finding comes after independent reports of a four-fold

Table 7 Standardized mortality ratios (SMR) for multiple myeloma in installation LAAs and relative risks (RR) of installation LAAs to control LAAs, 25-74 years age group, discrete distance zones

Grouping		Discrete distance zone				Sign [§]	Trend Probability
		1	2	3	4		
Pre-1955	SMR	90.3	109.7	104.5	108.3	—	0.017
	RR	0.79	1.14	1.12	1.14		
All CEBG	SMR	50.9	90.1	86.4	125.1	—	0.171
	RR	0.54	0.72	0.87	1.02		
Winfrith	SMR	147.3	—‡	90.1	62.5	+	0.223
	RR	1.77	—	0.71	0.46		
Sellafield	SMR	—‡	57.8	92.0	81.2	+	0.640
	RR	—	0.71	0.78	0.48		
All installations	SMR	88.5	100.9	101.5	106.0	—	0.093
	RR	0.79	0.99	1.06	0.99		

‡ No LAAs in this distance zone.

§ Sign of trend in RR with increasing proximity to installation.

|| Probability of trend arising due to chance (two-sided test).

Table 8 Standardized mortality ratios (SMR) for liver cancer, lung cancer and all malignancies in installation LAAs and relative risks (RR) of installation LAAs to control LAAs, 25–74 years age group, all CEBG installations, all distance zones, before and after start-up

Grouping		Cumulative distance zone							
		Before	1 After	Before	2 After	Before	3 After	Before	4 After
a Liver cancer									
	All CEBG								
	SMR	85.5	163.3	224.6†	189.0*	178.5†	148.8*	162.4†	136.0*
	RR	1.13	1.41	2.21	3.03*	1.91	1.93*	1.59	1.82*
b Lung cancer									
	All CEBG								
	SMR	110.3	107.8	102.4	100.7	96.6	97.2	92.2	95.0
	RR	1.06	1.04	1.03	1.02	1.04	1.06†	1.01	1.05†
c All malignancies									
	All CEBG								
	SMR	109.2	101.4	105.5	103.5†	99.1	99.1	97.9	98.4
	RR	1.05	0.95	1.06	1.02	1.02	1.03†	1.00	1.04*

* $P < 0.01$ (one-sided test for increase above 100 (SMR) or unity (RR)). † $P < 0.05$ (one-sided test for increase above 100 (SMR) or unity (RR)).

increase in leukaemia mortality in the age group 0–24 years in Millom Rural District, immediately south of Sellafield⁷. Other increases in childhood leukaemia incidence, based on registration data, have also been reported around two nuclear installations in Scotland (Dounreay¹⁵ and Hunterston¹⁶) and around Aldermaston¹⁷. However, current assessments of annual radiation doses, in conjunction with estimates of the risk of leukaemia per unit dose, indicate that the estimated doses received by populations living in the vicinity of nuclear installations are not sufficiently high to cause any detectable increase in leukaemia^{6,18}. A recent study of the effects of fallout from atmospheric nuclear weapons testing in the 1950s and 1960s provides good evidence that the estimate of the risk of leukaemia per unit dose is not much too low¹⁹. It follows that either the received doses are grossly underestimated or the excess risk of leukaemia must be due to some other factors that differ between the installation and control LAAs. Thus, on present knowledge, it is difficult to attribute the leukaemia results found here to ionizing radiation, but it is also unclear what the other factors are that distinguish the installation LAAs from the LAAs with which they have been compared.

The other positive observation in the age group 0–24 years, that for all brain cancer, seems most likely to be due to chance. The detailed results for this cancer type are contradictory. The increasing risk with proximity to Sellafield contrasts with a decreasing risk with proximity to the CEBG installations. There is no suggestion of a hazard round Winfrith or the pre-1955 installations. The one significantly raised RR in discrete zone 2 around Sellafield derives from a comparison between five and zero observed deaths, while in discrete zones 3 and 4 the corresponding relative risks are less than unity (0.86 and 0.00 respectively).

Of the positive results in the older age group, 25–74 years, those for Hodgkin's disease present the most consistent evidence for an association with proximity to the installations. There is a tendency for the relative risk to increase with

proximity to the installations around the pre-1955 installations, around Sellafield and Winfrith individually, and around all installations combined (Table 6). There is no notable trend in either direction around the CEBG installations. It should be noted, however, that the SMRs tend to be below or only slightly above the regional standard, except around Winfrith where an exceptionally high SMR was recorded in zone 1 (540.5 based on four cases).

Studies of populations exposed to high doses of radiation have not, so far, shown significant increases of Hodgkin's disease¹². However, the available data are not inconsistent with a small increase. It is also entirely possible that the present results for Hodgkin's disease could be a consequence of the possible socio-economic or environmental differences between installation and control LAAs, especially as there is some evidence that the causes of the disease are influenced by social factors²⁰. The positive result for all lymphomas in the innermost distance zone around the pre-1955 installations is accounted for by the result for Hodgkin's disease. If the latter is subtracted out then the remaining "other lymphomas" category shows no increased risk.

Previous observations on mortality from multiple myeloma are perhaps the most suggestive of all the data on adult cancers of a possible hazard associated with nuclear installations. The consistency of the results amongst both Sellafield⁹ and the Hanford^{10,11} workers has suggested that there may be a special hazard of myelomatosis associated with these installations. A dose-related increase in this disease has also been documented in Japanese atom-bomb survivors¹³. The observation in this study of a small increased relative risk in the LAAs of the coastal regions to the north and south of Sellafield (RR = 1.11, one-sided $P = 0.043$), and the above average value of 107.0 for the SMR in the same coastal areas provides some additional evidence for the association. The hypothesis would, however, seem to be contradicted by the tendency for the RRs to decrease with increasing proximity to all installations combined (two-sided $P = 0.093$) and par-

ticularly with increasing proximity to the pre-1955 installations (two-sided $P = 0.017$) and by the low SMRs in the LAAs immediately adjacent to Sellafield. The latter three LAAs are all on the coast and could be combined with the other 58 coastal LAAs. However, because they represent a very small proportion of the population in the group of coastal LAAs, the overall RR in the group is only marginally affected (RR = 1.09, one-sided $P = 0.079$).

All of the positive results around the CEBG installations are probably largely explained by the socio-economic differences between installation and control LAAs. This is made most apparent by considering the results for lung cancer and all malignancies in Tables 2 and 3. Both of these have RRs which are significantly elevated around the CEBG installations and significantly lowered around the pre-1955 installations and all installations combined. As lung cancer is such a major component of all malignancies, the similarity between the two categories is not surprising. Analysis of the results in discrete zones shows highly significant trends in RRs which decrease with increasing proximity to the installation for both lung cancer and all malignancies (two-sided $P = 0.005$ and 0.001 respectively around the pre-1955 installations and $P = 0.001$ and < 0.001 respectively around all installations combined). No significant trends are apparent around the grouped CEBG installations or around Sellafield or Winfrith.

The most likely explanation for these results is that social factors, especially smoking behaviour, have different distributions in the LAAs around the pre-1955 installations, especially within the inner distance zones, compared with the LAAs around the CEBG installations. The matching of control LAAs does not fully allow for this and, as such large numbers of deaths are involved, even moderate differences between installation LAAs and control LAAs produce statistically highly significant results.

A similar explanation can be made in relation to liver cancer which, like lung cancer, is only elevated around the CEBG

installations. Liver cancer has not, to our knowledge, been associated with radiation exposure at low doses. A large excess of liver tumours has been seen in patients injected with the radioactive contrast medium thorotrast¹², and the liver was selected as a specific site for this study because inhaled or ingested plutonium might act in the same way. However, the CEGB installations do not discharge substantial quantities of plutonium⁶ and one would have expected any effect to have been more evident in LAAs around the pre-1955 installations. The diagnosis of liver cancer as a cause of death after about 45 years of age is, moreover, extremely unreliable and most deaths attributed to this cause are likely to be due to metastases from primary cancers in the lung, stomach or large bowel. It is, therefore, only to be expected that liver cancer will show a similar relationship with vicinity to nuclear installations as is shown by lung cancer and all malignancies.

Further evidence that the significant results around the grouped CEGB installations are not caused by radioactive discharges from the installations is provided by comparing the RRs in the LAAs before and after start-up (Table 8). This shows very little difference between the two time periods in the SMRs or the RRs for liver cancer, lung cancer, or for all malignancies.

The result for "unspecified brain and central nervous system tumours" in the age group 25-74 years is presumably either due to chance or to diagnostic artefacts. The excess is more than compensated for by the deficiency of brain tumours specified as benign or malignant.

Summary

The OPCS report on cancer incidence and mortality in the vicinity of nuclear installations in England and Wales provides a mass of information that is so large that it should be possible to detect quite small changes in disease levels with considerable confidence. The data on cancer mortality are less subject to selective bias than the registration data on which incidence rates are based, and they provide the firmest grounds on which evidence of any effect can be obtained.

These data show conclusively that there has been no general increase in cancer mortality in the vicinity of nuclear installations in a 22-year period beginning several years after the opening of the installations that have released the largest amounts of radionuclides to the environment. On the contrary, the mortality from cancer has tended to be lower in the LAAs in the vicinity of nuclear installations than in control LAAs selected for their presumed comparability with the former. This is unlikely to be due to a protective effect of ionizing radiation and suggests that, despite the efforts that were made to choose

Table 9 Standardized mortality ratios (SMR) for lymphoid leukaemia in installation LAAs and relative risks (RR) of installation LAAs to control LAAs, individual pre-1955 installations, 0-24 years age group, discrete distance zones

Installation		Discrete distance zone				Trend	
		1	2	3	4	Sign [§]	Probability
BNFL Springfields	SMR	134.2	104.8	184.9*	170.9	+	0.530
	RR	2.35†	1.95	1.82†	1.44		
BNFL Capenhurst	SMR	70.5	116.4	58.4	81.7	+	0.811
	RR	1.12	0.84	0.28	1.09		
Amersham Int. plc	SMR	171.7	85.6	77.4	93.2	+	0.040
	RR	3.95†	0.87	1.02	0.83		
MOD, Aldermaston	SMR	—‡	159.1	89.4	46.9	+	0.701
	RR	—	0.90	0.77	0.52		
UKAEA, Harwell	SMR	77.6	67.8	67.1	—‡	+	0.238
	RR	2.54	0.37	0.41	—		

* $P < 0.01$ (one-sided test for increase above 100 (SMR) or unity (RR)).

† $P < 0.05$ (one-sided test for increase above 100 (SMR) or unity (RR)).

‡ No LAAs in this distance zone.

§ Sign of trend in RR with increasing proximity to installation.

|| Probability of trend arising due to chance (two-sided test).

comparable control areas, there were non-installation differences between the populations relevant to the risk of dying from one or other type of cancer.

Detailed examination of the few types of cancer that were relatively more common in the installation areas suggests that several of the differences were most likely to be due to chance, diagnostic artefacts or social factors rather than to any hazard specifically related to the installations. One disease provides a possible exception:

Appendix Sites of cancer, and combined groupings of sites*, which are covered in the OPCS report† and are discussed in this article

All malignant neoplasms (Mn)
 All leukaemia
 Lymphoid leukaemia†
 Non-lymphoid leukaemia†
 Myeloid leukaemia†
 Monocytic leukaemia†
 Other leukaemia†
 Unspecified leukaemia†
 Mn of liver
 Mn of lung‡
 Mn of bone
 Mn of testis
 Mn of thyroid
 All brain
 Mn of brain
 Benign neoplasms of brain and central nervous system‡
 Unspecified neoplasms of brain and central nervous system‡
 All lymphomas
 Hodgkin's disease§
 Other lymphomas§
 Multiple myeloma‡
 All lymphomas plus leukaemia

* Three of the grouped cancer sites examined in the OPCS report (namely, all malignancies less leukaemia, all malignancies less leukaemia and lung cancer, and all malignancies less leukaemia and brain cancer) were included for the sake of completion of specific analyses. We have excluded them from our analyses.

† Information on leukaemia subtypes was only available since 1968.

‡ For these cancer sites, for the age group 0-24 years only, data relating to the grouped CEGB installations have been used since 1966 and not since start-up of the installation.

§ For these cancer sites, data relating to the grouped CEGB installations have been used since 1966 and not since start-up of the installation for all age groups.

namely, leukaemia in the age group 0-24 years. Two other diseases need further investigation, multiple myeloma and Hodgkin's disease in the older age group 25-74 years. The excess mortality rates recorded from these cancers were not large, and it has yet to be established that they are not due to general confounding by other environmental or socio-economic factors.

We thank Miss Cynthia Bates for typing the numerous drafts of this paper.

1. Cook-Mozaffari, P. J., Ashwood, F. L., Vincent, T., Forman, D. & Alderson, M. *Cancer incidence and mortality in the vicinity of nuclear installations. England and Wales, 1950-1980* (Studies on Medical and Population Subjects, No. 51) (HMSO, London, 1987).
2. Office of Population Censuses and Surveys. *The Registrar General's Statistical Review of England and Wales for the year 1971* (HMSO, London, 1973).
3. Swerdlow, A. J. *J. R. Statist. Soc. A* **149**, 146-160 (1986).
4. Cook-Mozaffari, P. *Lancet* **i**, 855-856 (1987).
5. Office of Population Censuses and Surveys. *Occupational Mortality decennial supplement, 1970-1972, England and Wales*. (HMSO, London, 1978).
6. Clark, M. J. & Kelly, G. N. *Nucl. Energy* **21**, 275-288 (1982).
7. *Investigation of the possible increased incidence of cancer in west Cumbria*. (HMSO, London, 1984).
8. Committee on Medical Aspects of Radiation in the Environment (COMARE), First Report (HMSO, London, 1986).
9. Smith, P. G. & Douglas, A. J. *Br. med. J.* **ii**, 845-854 (1986).
10. Gilbert, E. S. & Marks, S. *Radiat. Res.* **79**, 122-148 (1979).
11. Tolley, H. D., Marks, S., Buchanan, J. A. & Gilbert, E. S. *Radiat. Res.* **95**, 211-213 (1983).
12. *The Effects on Populations of Exposure to Low Levels of Ionizing Radiation*. (Nat. Acad. Sci., Washington DC, 1980).
13. Kato, H. & Schull, W. J. *Radiat. Res.* **90**, 395-432 (1982).
14. Adams, N. *Phys. Med. Biol.* **26**, 1019-1034 (1981).
15. Heasman, M. A., Kemp, I. W., Urquhart, J. D. & Black, R. *Lancet* **i**, 266 (1986).
16. Heasman, M. A., Kemp, I. W., MacLaren, A., Trotter, P., Gillis, C. R. & Hole, D. J. *Lancet* **i**, 1188-1189 (1986).
17. Roman, E., Beral, V., Carpenter, L., Watson, A., Barton, C., Ryder, H. & Aston, D. L. *Br. med. J.* **294**, 597-602 (1987).
18. Hughes, J. S. & Roberts, G. N. *The radiation exposure of the UK population, 1984 Review: NRPB-R173* (HMSO, London, 1984).
19. Darby, S. C. & Doll, R. *Br. med. J.* **294**, 603-607 (1987).
20. Grufferman, S. in *Cancer, Epidemiology and Prevention* (eds Schottenfeld, D. & Fraumeni, J. F.) 739-753 (W. B. Saunders, Philadelphia, 1982).

The authors are at the Imperial Cancer Research Fund Cancer Epidemiology and Clinical Trials Unit, Radcliffe Infirmary, Oxford OX2 6HE, UK.

Structure of the human class I histocompatibility antigen, HLA-A2

P. J. Bjorkman^{*†‡}, M. A. Saper^{*}, B. Samraoui^{*}, W. S. Bennett^{*‡},
J. L. Strominger^{*} & D. C. Wiley^{*§}

^{*} Department of Biochemistry and Molecular Biology, Harvard University, Howard Hughes Medical Institute, Harvard University, Cambridge, Massachusetts 02138, USA

[†] Department of Medical Microbiology, Stanford University, Stanford, California 94305, USA

[‡] Present addresses: Department of Medical Microbiology, Stanford University, Stanford, California 94305, USA (P.J.B.) and Max-Planck-Institut für Molekulare Genetik, Abteilung Wittmann, D-1000 Berlin 33, FDR (W.S.B.)

[§] To whom correspondence should be addressed

The class I histocompatibility antigen from human cell membranes has two structural motifs: the membrane-proximal end of the glycoprotein contains two domains with immunoglobulin-folds that are paired in a novel manner, and the region distal from the membrane is a platform of eight antiparallel β -strands topped by α -helices. A large groove between the α -helices provides a binding site for processed foreign antigens. An unknown 'antigen' is found in this site in crystals of purified HLA-A2.

HLA (human leukocyte antigen) molecules are polymorphic membrane glycoproteins found on the surface of nearly all cells. Multiple genetic loci within the major histocompatibility complex (MHC) encode these proteins, and one individual simultaneously expresses several polymorphic forms from a large pool of alleles in the population. HLA molecules (also known as class I histocompatibility antigens) are the targets of antibodies and cytotoxic T lymphocytes (CTL) during rejection of foreign transplants^{1,2}. They are also recognized by T cells together with viral antigens on infected cell surfaces, a phenomenon known as MHC restricted recognition³. In contrast to antibodies, that can bind to free virus or soluble antigens, T-cell receptors only recognize foreign antigens that are associ-

ated with a particular HLA molecule. Because one individual expresses a limited set of different HLA molecules, a central question has been how these few HLA molecules can interact with so many foreign antigens. Limitations in the ability of a particular HLA molecule to associate with all antigens may explain the linkage of histocompatibility antigens to variations in susceptibility to human diseases⁴, and the immune system's responsiveness to particular antigens⁵.

Recent work has shown that virus-specific CTL will lyse an uninfected target cell of appropriate class I specificity to which peptide fragments of a viral protein have been added⁶. T-helper cells had previously been shown to recognize fragments of protein antigens in association with class II histocompatibility

Fig. 1 Representative portions of the HLA-A2 model superimposed on the 3.5 Å averaged electron density map. Electron density is shown for residues 66–77 in the α_1 long helix (a) and for residues 98–103 in β -strand 1 of α_2 (b). Contours are drawn at 0.9 σ . Rotation and translation parameters were determined that relate the monoclinic and orthorhombic maps, and these maps were iteratively averaged³⁴. The input P2₁2₁ map was calculated with phases from a single derivative, and the input P2₁ map was calculated with phases from four derivatives combined with phases from a partial model of the structure (model did not include side chains in α_1 and α_2 ; see text for details). As a control against bias from the partial model included in the P2₁ map, the real space averaging was repeated starting with the model-free, uninterpretable monoclinic and orthorhombic maps, resulting in a similar and interpretable averaged map. It is possible that some side chains located on the surface of the molecule or near crystal contacts have different conformations in the two crystal forms. The final positioning of side chains awaits extension to higher resolution and refinement. (Current *R* factor between P2₁ calculated and observed amplitudes is 26.2% with good geometry after 24 cycles of 6–3.2 Å refinement using the program CORELS³³.)

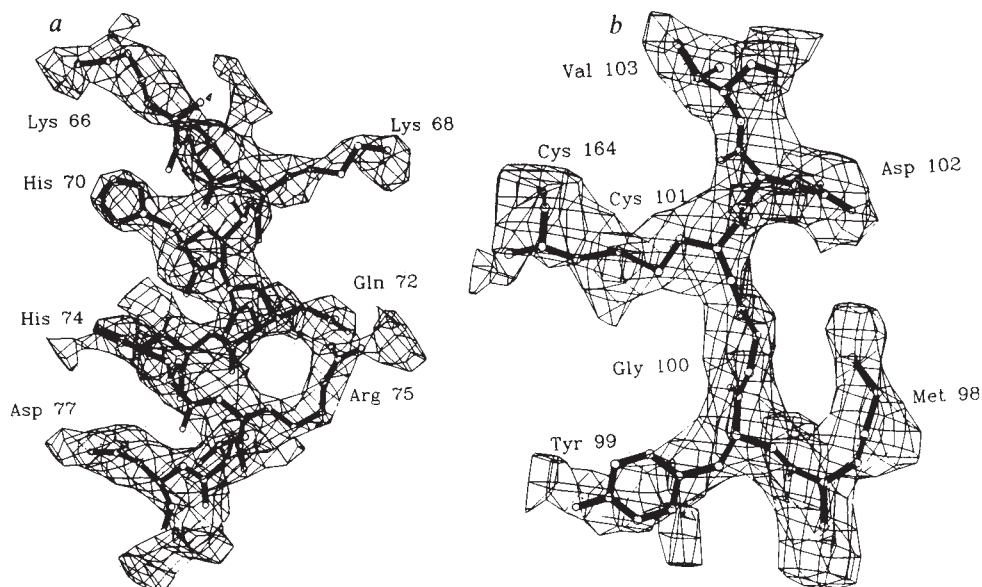


Table 1 Data collection, phasing and map averaging statistics

Mean figure of merit: Monoclinic – 0.60 for 5,759 reflections (3.5 Å resolution) Orthorhombic – 0.40 for 4,549 reflections						
Derivative	d_{lim} (Å)	Number of films (source)	Number of measurements	Number of reflections, (% complete)	$R^{\dagger}$	r.m.s. $f_h/E^{\ddagger}$ (15–3.5 Å)
P2 ₁ ($a = 60.4$ Å, $b = 80.4$ Å, $c = 56.5$ Å, $\beta = 120.4^\circ$)						
Native	2.7	120 (DESY/CHESS)	80,891	13,513 (96%)	0.101	—
K ₂ PtCl ₄	2.7	44 (DESY)	42,299	12,808 (89%)	0.095	1.78
HgI ₂	2.7	40 (DESY)	37,101	12,696 (81%)	0.105	0.92
K ₂ OsCl ₆	2.7	35 (DESY)	30,888	12,110 (78%)	0.091	0.52
PCMP*	2.8	54 (SSRL/CHESS)	34,295	10,247 (83%)	0.114	0.97
P2 ₁ 2 ₁ 2 ₁ ($a = 60.2$ Å, $b = 80.4$ Å, $c = 112.2$ Å)						
Native	2.8	24 (CHESS/XEN)	30,590	10,638 (70%)	0.087	—
K ₂ PtCl ₄	3.5	— (XEN)	15,106	5,028 (70%)	0.094	1.45
Map averaging statistics at 3.5 Å resolution						
	$R^{\S}$		r.m.s. phase difference [¶] (degrees)			
	P2 ₁	P2 ₁ 2 ₁ 2 ₁	P2 ₁	P2 ₁ 2 ₁ 2 ₁		
Cycle 1	0.342	0.343	47	38		
Cycle 5	0.247	0.255	68	78		
Cycle 10	0.166	0.185	73	83		

Film data were collected on 5° oscillation photographs (P2₁) or 2–3° oscillation photographs (P2₁2₁2₁) using synchrotron X-ray sources, and processed and scaled as described²⁸. The Xentronics area detector data were recorded as 5° oscillation frames, then merged into batches of 20 frames and scaled to the native orthorhombic film data. The monoclinic derivatives used in phasing were chosen after screening over 100 heavy atom compounds, and recording complete film data sets from 13 different heavy atom derivatives, 9 of which did not prove useful in phasing. The most useful monoclinic derivative (K₂PtCl₄) was the only dataset with an overall phasing power (r.m.s. f_h/E) greater than 1.0 to 3.5 Å resolution. The other derivative datasets had phasing powers greater than 1.0 only at lower resolution (~6.0 Å). The K₂PtCl₄ orthorhombic data were truncated at 3.5 Å resolution due to poor quality at higher resolution, and these data determine the resolution limit of the averaged electron density maps (Fig. 1a,b). Statistics are presented for 10 cycles of real-space averaging of the monoclinic and orthorhombic electron density maps.

* *p*-Chloromercuriphenol; † $R_I = \sum |I - \langle I \rangle| / \sum \langle I \rangle$; ‡ f_h , heavy atom structure factor; E is the residual lack of closure; || DESY, Deutsches Elektronen Synchrotron; CHESS, Cornell High Energy Synchrotron Source; SSRL, Stanford Synchrotron Radiation Laboratory; XEN, Xentronics Area Detector.

§ R , crystallographic R -factor between observed structure factors and calculated structure factors from the averaged map.

¶ r.m.s. phase difference between calculated phases and phases used in the starting map.

molecules⁷. Short synthetic peptides (10–20 residues long) have also been shown to bind to purified class II proteins^{8,9} at what appears to be a single binding site¹⁰. By analogy, it is likely that the homologous class I molecules bind antigenic peptides, and that the HLA-peptide complex is recognized by T-cell receptors on CTL¹¹.

Class I molecules (HLA-A, B, C in humans, H-2K, D, L in mice) are composed of two polypeptide chains¹²; a heavy chain (relative molecular mass 44,000; M_r 44K) which spans the membrane bilayer, and the non-covalently attached light chain, β_2 -microglobulin (β_2m —12K)^{13,14}. The extracellular portion of the heavy chain is divided into three domains, α_1 , α_2 and α_3 (ref. 15), each ~90 amino acids long and encoded on separate exons¹⁶. The α_3 domain and β_2m are relatively conserved and show amino-acid sequence homology to immunoglobulin constant domains^{15,17–20}. The polymorphic α_1 and α_2 domains show no significant sequence homology to immunoglobulin constant or variable domains, but have been reported to show weak sequence homology to each other²¹. We have crystallized a soluble fragment of HLA-A2 (ref. 22), composed of α_1 , α_2 , α_3 and β_2m , after removing the transmembrane anchor of the heavy chain by papain digestion^{23,24}. (HLA-A2, formerly named Mac, was the first polymorphic human leukocyte antigen identified²⁵.)

The structure determination of HLA-A2 was undertaken to gain an understanding of where the polymorphic residues are located on the structure, how and where antibodies and T-cell receptors recognize the molecule, and how the molecule interacts with foreign antigens. Here we report a description of HLA-A2 from a 3.5-Å X-ray crystallographic structure determination. The membrane-proximal α_3 and β_2m domains have tertiary structures resembling antibody domains, but are paired by a novel interaction not previously seen in immunoglobulin struc-

tures. The α_1 and α_2 domains are nearly identical to each other in structure, and are not similar to immunoglobulin constant or variable domains. The domains α_1 and α_2 form a platform composed of a single β -pleated sheet topped by α -helices with a long groove between the helices. Electron density which is not a part of the HLA molecule is observed in this site between the helices, presumably representing an unknown bound antigen. The location of polymorphic residues, antibody binding sites, and mutations affecting recognition by CTL are analysed in the accompanying article, which defines the binding site for processed foreign antigens²⁶.

Structure determination

Soluble HLA-A2 was purified after papain digestion of plasma membranes from the homozygous human lymphoblastoid cell line JY^{23,24}. Papain cleaves the HLA heavy chain at residue 271, thirteen residues from the transmembrane region²⁷, yielding a molecule composed of α_1 , α_2 , α_3 and β_2m . When purified, 3–4 mg protein is obtained from 200 litres of human tissue culture cells.

The crystallization and a proposed packing model for the two crystal forms of HLA-A2 and HLA-A28 have been described previously^{22,28}. HLA-A2 crystallizes in two space groups: monoclinic P2₁ and orthorhombic P2₁2₁2₁. Both crystal forms grow as very thin (20 μ m) plates ~0.5 mm square. Occasional 100 μ m thick orthorhombic crystals have been grown. Data from the 20 μ m thick monoclinic crystals were collected using synchrotron X-ray sources. Part of the data from the orthorhombic crystals were collected using a Xentronics area detector^{29,30}. Data collection and phasing statistics are presented in Table 1.

A monoclinic electron density map was calculated to 3.5 Å resolution using isomorphous differences and anomalous scat-

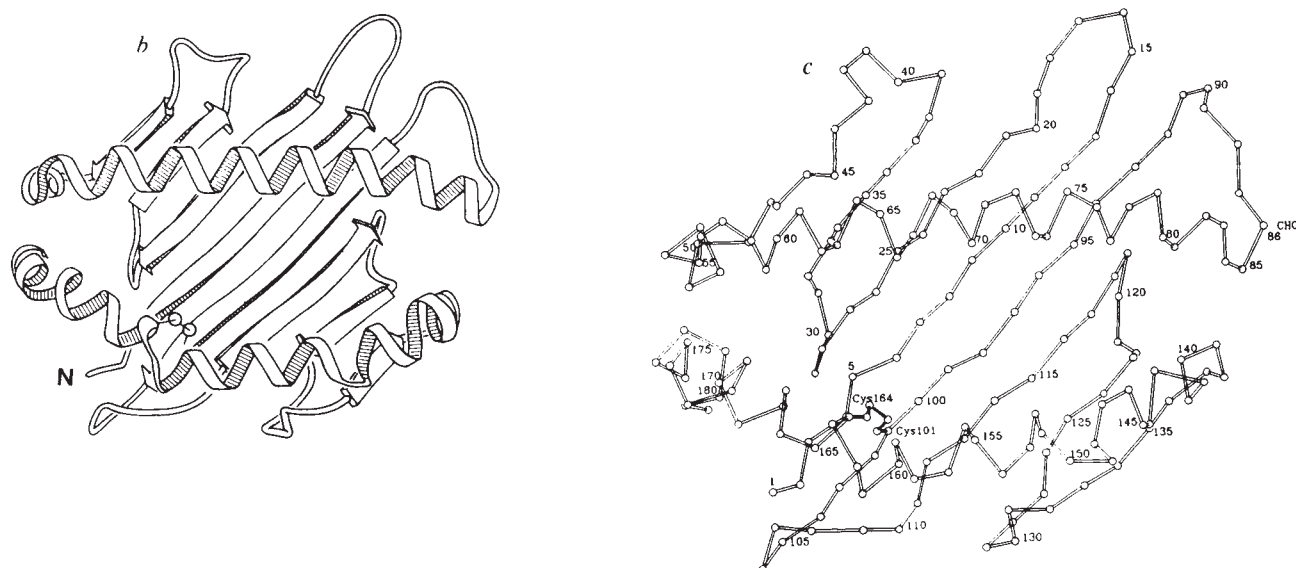


Fig. 2 Schematic representation of the structure of HLA-A2. The β -strands are shown as thick arrows in the amino to carboxy direction, α -helices are represented as helical ribbons. Connecting loops are depicted as thin lines. Disulphide bonds are indicated as two connected spheres. *a* (Facing page), schematic representation of the four domains of HLA-A2. The molecule is shown with the membrane proximal immunoglobulin-like domains (α_3 , β_{2m}) at the bottom, and the polymorphic α_1 and α_2 domains at the top. The indicated C-terminus of α_3 is the papain cleavage site. In the membrane-bound molecule, another 13 amino acids extend past the cleavage site toward the membrane, which we assume to be horizontal at the bottom of this picture. This orientation places the helices of the α_1 and α_2 domains, and the antigen recognition site located between them, on the top surface of the molecule. The domains α_1 and α_2 form a platform with a single eight-stranded β -pleated sheet (seen edge on), covered by α -helices. *b*, Schematic representation of the top surface of HLA-A2. The α_1 and α_2 domains are shown as viewed from the top of the molecule, showing the surface that is presumably contacted by a T-cell receptor (90° from Fig. 2*a*). The N terminus of α_1 is indicated. Each domain consists of four antiparallel β -strands followed by a long helical region, and the domains pair to form a single eight-stranded β -sheet topped by α -helices. The disulphide bond connects residue 164 in the α_2 long helix to residue 101 in the first β -strand of α_2 . The residue analogous to Cys 101 in α_1 is Ser 11; it is 8 Å from Ser 71 (C_α to C_α) on the α_2 long helix. *c*, α -carbon backbone of the α_1 and α_2 domains. The domains are shown in the same orientation as in *b*. Every fifth residue and the oligosaccharide attachment site (86) are labelled.

Fig. 3 (Right) Schematic representation of the secondary structure of HLA-A2. The four β -strands and helical regions in α_1 occupy equivalent locations as the β -strands and α -helices in α_2 , and the two domains are folded into homologous tertiary structures (Fig. 4). In α_3 and β_{2m} , the seven β -strands fold into structures resembling immunoglobulin constant domains. The positions of cysteine residues involved in disulphide bonds, and the location of the glycosylation site at residue 86 are indicated.

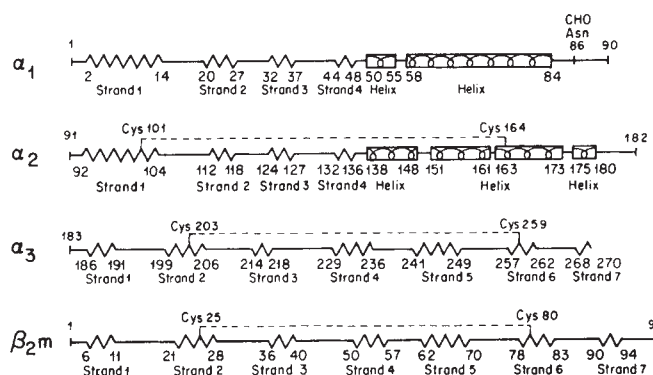


Fig. 4 (Facing page) Comparison of the structures of the α_1 and α_2 domains. *a*, The α -carbon backbone of α_1 . N and C termini are labelled. *b*, The α -carbon backbone of α_2 . N and C termini are labelled. *c*, The α_1 (blue) and α_2 (red) domains as they are paired in the HLA-A2 structure. The N-terminal β -strand in each domain is depicted with main-chain atoms to emphasize the hydrogen bonds (yellow lines) between these strands that form a single, eight-stranded β -pleated sheet from the four-stranded sheets in each domain. The rest of each domain is shown as the C_α backbone. Yellow stars highlight the C_α atom of any residue with a main or side chain atom within 4 Å of an atom located in the partner domain (see Table 2). The N terminus of α_1 is labelled. *d*, The α -carbon backbones of α_1 (blue) and α_2 (red) superimposed to show structural similarity; α_1 and α_2 were aligned after a least-squares fit of the α -carbon coordinates of the β -strand and α -helical residues in each domain. The average r.m.s. positional difference is 3.1 Å.

Fig. 5 (Facing page) Location of the pseudo-symmetry axes in the HLA-A2 molecule. Domains α_1 (red) and α_2 (yellow) are related by an approximate dyad axis of symmetry (orange arrow). (The relationship is a rotation of $\sim 177^\circ$ followed by a 0.7 Å translation.) Domains α_3 (blue) and β_{2m} (purple) are related by a 146° rotation (blue arrow) followed by a 13-Å translation. The two axes do not intersect, and deviate from colinearity by 25° (in projection).

Fig. 6 (Facing page) Van der Waals surface representation⁵⁴ of the top of the HLA-A2 molecule (*a*) showing the deep groove identified as the antigen recognition site²⁶ and the electron density (*b*) found in this site in crystals of HLA-A2. The surface was generated with the program GRIDCTR written by Mark Handschumacher and Fred Richards. The molecule is shown from the top with the C_α backbone (pink) oriented as in Fig. 2*b,c*. *a*, Van der Waals surface of the α_1 and α_2 domains (blue) showing a deep groove running between the helical regions in the α_1 and α_2 domains. This groove has been identified as the recognition site for processed antigens (see text and ref. 26 for details). *b*, The extra electron density (red contours at 0.8σ) found in both crystal forms of HLA-A2 is shown superimposed on the Van der Waals surface of α_1 and α_2 (blue). The extra density represents a bound molecule or mixture of molecules that co-purified and co-crystallized with HLA-A2. Some regions of the extra density may be wide enough to accommodate a peptide in an α -helical conformation, but in the absence of the knowledge of its composition, or whether it is one species or a mixture, it is not possible to interpret its structure unambiguously at the current resolution of the electron density maps (3.5 Å).

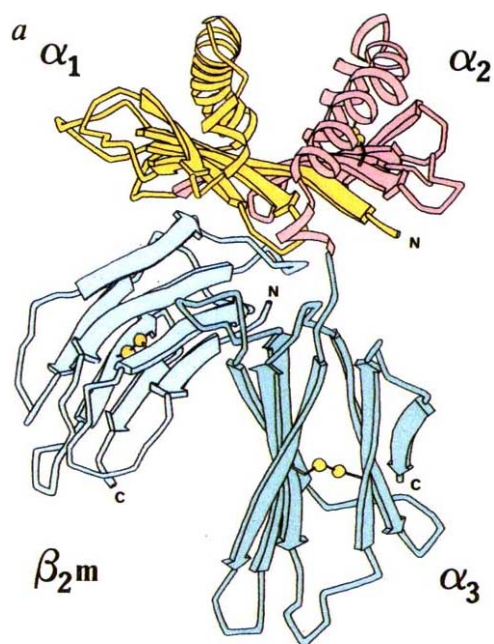


Fig. 2a

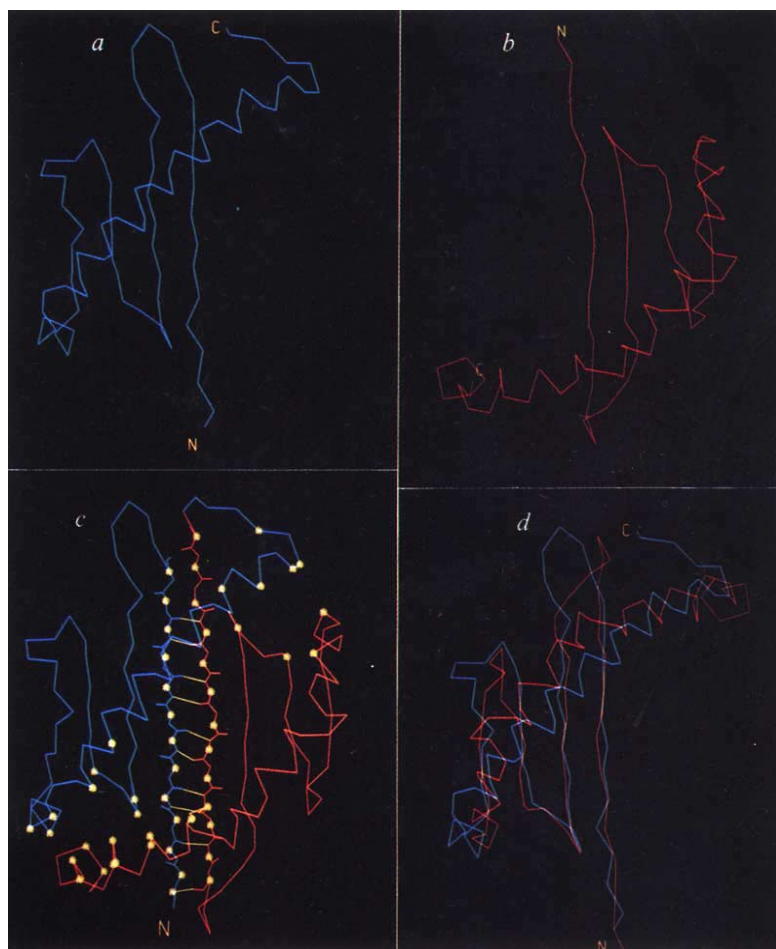


Fig. 4

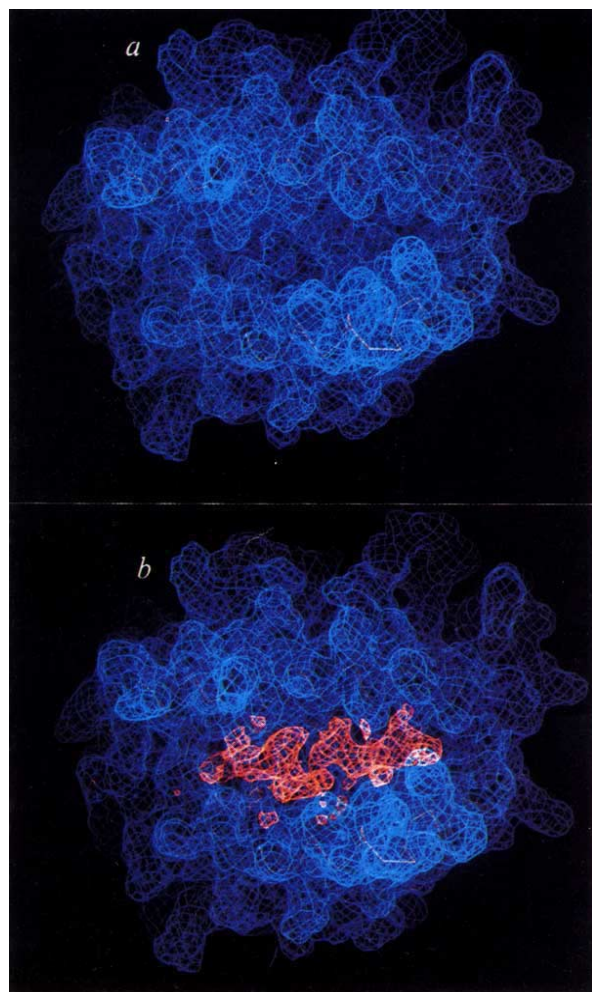


Fig. 6

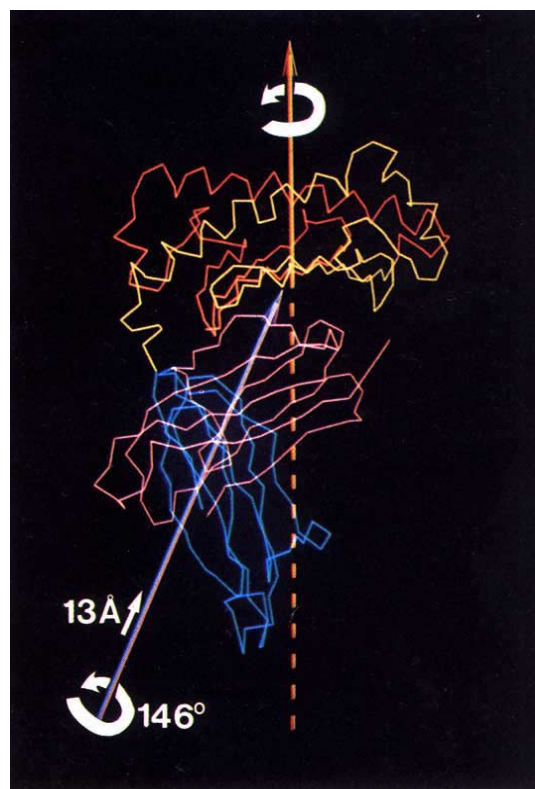


Fig. 5

tering from four heavy atom derivatives and modified by iterative solvent flattening³¹. Due to the poor quality of the heavy atom derivatives, the map was not fully interpretable, although the molecular outline and the two immunoglobulin-like domains of α_3 and β_2m were evident. Subsequent electron density maps derived from many cycles of model building, phase combination³² and CORELS refinement³³ permitted the complete structures of α_3 and β_2m to be determined, and ~80% of the main chain of α_1 and α_2 to be fit as segments of polyalanine. To complete tracing the polypeptide chain in α_1 and α_2 , a single isomorphous replacement electron density map was calculated from 3.5 Å data from orthorhombic crystals. The relationship between the two crystal forms was established by a six-dimensional real-space search using the monoclinic partial model, and confirmed by the location of a common heavy atom derivative. The monoclinic and orthorhombic maps were averaged inside their molecular envelopes using the Bricogne algorithms³⁴. Solvent region densities were replaced by their average value, and new phases were calculated by Fourier transformation of the maps. Ten cycles of averaging in both crystal forms produced significant phase changes and *R*-factor improvements (Table 1), allowing the complete polypeptide chain to be traced unambiguously with the aid of the amino-acid sequence³⁵ (details will be published elsewhere; Saper *et al.*, manuscript in preparation). Despite the limited resolution (3.5 Å), 80% of the side chains are fitted in well-defined density (Fig. 1). In the future, least squares refinement using 2.7 Å native data in either space group will allow a more precise positioning of side chains.

Description of the structure

HLA-A2 consists of two pairs of structurally similar domains: α_1 has the same tertiary fold as α_2 , while α_3 has the same tertiary fold as β_2m . As oriented in Fig. 2a, the α_3 and β_2m domains are closest to the bottom of the figure. These domains are proximal to the cell membrane, which would probably be horizontal in this orientation (see Fig. 2 legend). Domains α_1 and α_2 form the top of the molecule with their helical sides facing away from the cell. A view of the 'top' of HLA-A2 (Fig. 2b) shows the surface that is presumably recognized by T cells. The cross-section of the top of the molecule is about 50 Å × 40 Å, and the molecule is about 70 Å in length.

Structures of α_3 and β_2m

The α_3 and β_2m domains are both β -sandwich structures composed of two antiparallel β -pleated sheets, one with four β -strands and one with three β -strands, connected by a disulphide bond. This tertiary structure has been described for constant regions of immunoglobulin molecules (for review, see refs 36, 37) and was expected from the significant sequence homology between α_3 , β_2m and constant regions^{15,17-20}. The structure of human β_2m as bound to the HLA-A2 heavy chain also appears similar to the structure of the free monomer of bovine β_2m ³⁸. A least-squares comparison between the α -carbon coordinates of the β -strand regions of α_3 with β_2m indicates an average r.m.s. positional difference of 1.4 Å. A similar comparison of α_3 and β_2m to an immunoglobulin constant region (C_{H3} from human Fc) yields an r.m.s. difference of 1.6 Å in each case. Similar comparisons among antibody constant domains give r.m.s. differences of 0.6–1.1 Å³⁸.

Structures of α_1 and α_2

The α_1 and α_2 domains each consist of an antiparallel β -pleated sheet spanned by a long α -helical region that is C-terminal to the four β -strands in the sheet (Figs 3, 4a,b). In the helical region of α_1 , a short α -helix (50–55) precedes a longer curved α -helix (58–84). These helices are almost at right angles (~110°). In α_2 , a short helix (138–148) precedes a longer helix (151–173) at an angle of ~130°. The long helix in α_2 is kinked at residue 162. An additional short helix in α_2 (177–181) connects α_2 to

α_3 . A disulphide bond in α_2 connects residue 101 in the N-terminal β -strand to the long helix at residue 164. In Fig. 4d, α_1 is superimposed on α_2 to show their structural similarity (overall r.m.s. difference 3.1 Å). Published secondary structure predictions^{39,40} can be compared with Fig. 3.

Interdomain contacts

The α_3 and β_2m domains contact each other in an interaction not found between pairs of constant domains in the known antibody structures (for example, C_{H1} and C_L in an Fab, the C_{H3} dimer in Fc). Antibody constant domains are related by a nearly exact dyad (180°) symmetry axis, with their four-stranded β -sheets forming the contact interface^{36,37}. In HLA-A2, although the immunoglobulin-like domains contact each other with their four-stranded sheets (Table 2), they are related by a 146° rotation followed by a 13 Å translation (Fig. 5).

The structurally similar α_1 and α_2 domains (Fig. 4a,b) are paired in the HLA molecule by an approximate dyad axis of symmetry, such that the four β -strands from each domain form a single antiparallel β -sheet with eight strands (Fig. 4c). This β -sheet is topped by the helical regions from each domain, which are separated by about 18 Å (centre to centre). The interface between α_1 and α_2 includes hydrogen bonding between the N-terminal β -strands of each domain at the centre of the eight-stranded sheet, and contacts between the C-terminal ends of the helical regions and β -sheet residues in the other domain (Fig. 4c, Table 2). The presence of approximate dyad symmetry has previously been observed between homologous domains in a single polypeptide chain, and is termed 'intramolecular dimerization'⁴¹. The particular dimeric interaction seen between α_1 and α_2 , involving the creation of a single β -sheet from two domains, has also been observed in a number of intermolecular dimers (concanavalin A⁴², aspartate carbamoyltransferase R chains⁴³, alcohol dehydrogenase⁵⁵), indicating that the α_1 - α_2 domain interface could be preserved in an intermolecular dimer such as that in class II histocompatibility antigens (see below).

Pseudo-symmetry axes, domain interfaces

The screw axis relating the two immunoglobulin-like domains of HLA-A2 (146° rotation and 13 Å translation) and the approximate dyad axis relating α_1 and α_2 are neither co-linear nor intersecting. In projection, the two axes deviate by 25° from co-linearity (Fig. 5). The presence of a local dyad axis had been anticipated by earlier crystallographic work^{22,28}, but the relationship between the immunoglobulin-like domains in HLA was unexpected. As this arrangement of domains is found in both crystal forms of HLA-A2, it is not likely to result from crystal-packing forces. Table 2 lists the residues in each domain that are within 4 Å of another domain. The residues in contact across the α_3 - β_2m interface are in the four-stranded β -sheets, but the contact pairs are not symmetrical as would be expected if the domains were related by a dyad. The β_2m domain also interacts with the central β -strands and loops of the α_1 - α_2 β -sheet. The interaction of β_2m with all three domains of the heavy chain of HLA was suggested by immunological⁴⁴ and circular dichroism data⁴⁵, and results in a molecule that is shaped quite differently from an Fab^{36,37}.

Carbohydrate

Both human and mouse class I histocompatibility antigens have an N-linked, complex oligosaccharide attached at Asn 86 (ref. 46). This residue is in a loop connecting α_1 to α_2 . Carbohydrate is visible in the HLA electron-density map extending away from the protein structure, but analysis of its structure awaits high resolution refinement. Several studies have demonstrated that glycosylation at Asn 86 is not necessary for the association of β_2m with the HLA heavy chain or cell-surface expression^{47,48}. In murine class I antigens there is a second glycosylation site at Asn 176, and H-2L^d, H-2D^b and H-2K^d molecules have a third at Asn 256 (ref. 49). Residue 176 is located in α_2 at a

Table 2 Inter-domain contacts

α_1	α_2	β_2	α_3
1	104	8	231, 244
2	104	10	234, 242
3	103, 168, 172, 180	11	242
4	101, 102, 164, 168	12	238
5	99, 100, 101, 164, 168, 171	13	188, 206
6	98, 99	22	237
7	98, 99	26	233, 234
8	97	65	237
9	95, 96, 97, 99	97	204
10	95, 96		
11	94, 95	β_2	α_1
12	93	33	12
13	92	53	25, 33, 35
28	171, 179	55	27
29	179	56	8, 9, 10
32	171	62	10
51	174, 175	63	6, 27
54	174		
55	170		
59	167, 171	β_2	α_2
63	171	1	119, 120
70	99	31	94, 96, 119
74	95, 97	56	96, 97
78	95	60	96, 115, 116, 117, 121
81	118, 123	62	96
84	123, 139, 142, 143		
85	118		
87	118	α_3	α_1
		209	29, 30
		211	30
		234	27, 33
		α_3	α_2
		181	176, 177, 179
		182	180
		209	179

All residues (main chain or side chain) of one domain within 4 Å of the second domain are listed. This table is included as a guide for interpreting mutagenesis and exon shuffling experiments that exchange domains between different histocompatibility molecules. The contact list may require minor updating after crystallographic refinement. The residue numbers can be assigned to their secondary structural location on the HLA-A2 structure using Fig. 3. The α_1 - α_2 contacts are shown on the structure of these domains in Fig. 4c.

position structurally homologous to Asn 86 in α_1 . The presence of carbohydrate at position 176 would increase the approximate symmetric nature of the α_1 - α_2 'intramolecular dimer'. Residue 256 is located on the loop between strands 5 and 6 in α_3 (Fig. 3) and is accessible to solvent in HLA-A2.

Antigen recognition site

A deep groove ~25 Å long and 10 Å wide runs between the two long α -helices of the α_1 and α_2 domains (Fig. 6a). The sides of the groove are formed by side chains from these helices, and the bottom is formed by side chains from the central β -strands of the α_1 - α_2 β -sheet. The groove is located on the top surface of the molecule (as oriented in Fig. 2a), and is therefore a likely candidate for the binding site for the foreign antigen that is recognized together with HLA by a T-cell receptor. The dimensions of the site are consistent with the expectation that class I molecules bind a processed antigen, probably a peptide. The site is lined with both polar and non-polar side chains. Some of these residues have been identified to be critical for T-cell recognition, as described more fully in an accompanying article²⁶.

In this site, we observe a large continuous region of electron density that is not accounted for by the polypeptide chain of the HLA molecule (Fig. 6b). This extra density is found in

electron density maps calculated in both space groups. The density level is comparable to the density level of the protein, suggesting that most or all of the crystalline HLA molecules have a molecule(s) bound in the proposed antigen binding site. It seems likely that the extra density is the image of a peptide or mixture of peptides that co-purified and co-crystallized with HLA-A2. Potential sources of such peptides are: (1) protein fragments produced during papain cleavage; (2) peptides from the fetal calf serum used in the growth of JY cells; (3) processed antigenic peptides from Epstein Barr virus used in the transformation of JY cells⁵⁰; and/or (4) endogenous 'self' peptides created during normal recycling and degradation of cellular proteins. In the absence of the knowledge of the composition of the bound molecule(s) and whether it is one species or a mixture, it is not possible to interpret its structure unambiguously at the current resolution (3.5 Å) of the electron density maps.

Other histocompatibility antigens

The structure of HLA-A2 can be used as a starting point for modelling other class I and class II histocompatibility antigens. Using the extensive sequence homology between murine and human class I molecules, it is possible to locate residues critical for murine CTL and serological epitopes on the HLA-A2 structure (see accompanying article)²⁶.

Class II molecules have a domain structure similar to class I, but the four domains of the class II chains are arranged on two polypeptide chains of roughly equal size that span the membrane bilayer (reviewed in ref. 51). The membrane-proximal domains of each chain are immunoglobulin-like with homology to α_3 and β_2 m. The N-terminal domains of the class II chains presumably contain the binding site for antigenic peptides, and one of these domains has weak sequence homology to class I α_1 and α_2 (ref. 52). Two features of the HLA-A2 molecule suggest that class II molecules could adopt a similar structure: (1) The 'intramolecular dimerization' between the α_1 and α_2 domains resulting in an edge-to-edge β -sheet contact is similar to inter-molecular dimer contacts observed in multisubunit molecules. It is therefore possible that class II N-terminal domains can form an intermolecular dimer with a structure similar to the α_1 and α_2 region of HLA-A2, thereby creating a binding site for antigens similar to that in HLA-A2. Residues in the first half of the membrane-distal domains of class II molecules have recently been shown to be critical for chain association⁵³, a result consistent with the proximity of the N-terminal β -strands of the class I α_1 and α_2 domains shown in Fig. 4c. (2) In the HLA-A2 structure, the N-terminus of β_2 m is 11 Å from the C-terminus of the α_1 domain. By replacing the loop that connects α_1 and α_2 (residues 85-93) with an extended chain connecting the end of α_1 to the N-terminus of β_2 m, a four-domain structure can be constructed with two domains per polypeptide chain that is similar to HLA-A2 (J. Brown *et al.*, manuscript in preparation).

These observations indicate that the HLA-A2 structure may be a good first-order model for interpreting the locations of polymorphic and functional residues on many class I and class II histocompatibility antigens.

We thank Anastasia Haykov for excellent technical assistance; Hans Bartunik and Klaus Bartels, Keith Moffat and Wilfried Schildkamp and Paul Phizackerley for help with synchrotron data collection, Mike Silver and Tom Garrett for help with the figures and our colleagues in the structural molecular biology group. We also thank Drs Dean Mann (NIH), D. Michael Strong and James Woody (U.S. Naval Research Unit) and Don Giard (MIT Cell Culture Facility) for provision of cells which made this work possible. B.S. thanks the Algerian Ministère de l'Enseignement et de la Recherche Scientifique for a postdoctoral leave. P.J.B. held an American Cancer Society postdoctoral fellowship during part of the work. The research was supported by the NIH and the Howard Hughes Medical Institute.

Note added in proof: Alpha-carbon coordinates are being deposited in the Brookhaven Data Bank.

Received 17 August; accepted 26 August, 1987.

1. Klein, J. in *Biology of the Mouse Histocompatibility Complex* (Springer, Berlin, 1975).
2. Bach, F. H. & van Rood, J. J. *New Engl. J. Med.* **295**, 806-813 (1976).
3. Zinkernagel, R. M. & Doherty, P. C. *Nature* **248**, 701-702 (1974).
4. Zinkernagel, R. M. *A. Rev. Microbiol.* **33**, 201-213 (1979).
5. McDevitt, H. O. & Tyan, M. L. *J. exp. Med.* **128**, 1-11 (1968).
6. Townsend, A. R. M. *et al. Cell* **44**, 959-968 (1986).
7. Shimonkevitz, R., Kappler, J., Marrack, P. & Grey, H. *J. exp. Med.* **158**, 303-316 (1983).
8. Babbitt, B., Allen, G., Matsueda, E., Haber, E. & Unanue, E. *Nature* **317**, 359-361 (1985).
9. Buus, S. *et al. Proc. natn. Acad. Sci. U.S.A.* **83**, 3968-2971 (1986).
10. Guillet, J. G., Laim-Z., Briner, T. J., Smith, J. A. & Geftter, M. L. *Nature* **324**, 260-263 (1986).
11. Forman, J. *Adv. Immun.* (in the press).
12. Cresswell, P., Turner, M. J. & Strominger, J. L. *Proc. natn. Acad. Sci. U.S.A.* **70**, 1503-1507 (1973).
13. Grey, A. M. *et al. J. exp. Med.* **138**, 1608-1612 (1973).
14. Peterson, P. A., Rask, L. & Lindblom, J. B. *Proc. natn. Acad. Sci. U.S.A.* **71**, 35-39 (1974).
15. Orr, H. T., Lopez de Castro, J. A., Lancet, D. & Strominger, J. *Proc. natn. Acad. Sci. U.S.A.* **76**, 5839-5842 (1979).
16. Malissen, M., Malissen, B. & Jordan, B. R. *Proc. natn. Acad. Sci. U.S.A.* **79**, 893-897 (1982).
17. Tragardh, L., Rask, L., Wiman, K., Fohlman, J. & Peterson, P. A. *Proc. natn. Acad. Sci. U.S.A.* **76**, 5839-5842 (1979).
18. Peterson, P. A., Cunningham, B. A., Berggard, I. & Edelman, G. M. *Proc. natn. Acad. Sci. U.S.A.* **69**, 1697-1701 (1972).
19. Smithies, O. & Poulik, M. D. *Science* **175**, 187-189 (1972).
20. Orr, H. T., Lancet, D., Robb, R. J., Lopez de Castro, J. A. & Strominger, J. L. *Nature* **282**, 266-270 (1979).
21. Orr, H. T., Lopez de Castro, J. A., Lancet, D. & Strominger, J. L. *Biochemistry* **18**, 5711-5719 (1979).
22. Bjorkman, P. J., Strominger, J. L. & Wiley, D. C. *J. molec. Biol.* **186**, 205-210 (1985).
23. Turner, M. J. *et al. J. biol. Chem.* **250**, 4512-4519 (1975).
24. Parham, P., Alpert, B. N., Orr, H. T. & Strominger, J. L. *J. biol. Chem.* **252**, 7555-7567 (1977).
25. Dausset, J. *Acta haematol.* **20**, 156 (1958).
26. Bjorkman, P. J. *et al. Nature* **329**, 512-518 (1987).
27. Springer, T. A. & Strominger, J. L. *Proc. natn. Acad. Sci. U.S.A.* **73**, 2481-2485 (1976).
28. Bjorkman, P. J. thesis, Harvard University (1984).
29. Durbin, R. *et al. Science* **232**, 1127-1132 (1986).
30. Blum, M., Metcalf, P., Harrison, S. C. & Wiley, D. C. *J. appl. Cryst.* **20**, 235-242 (1987).
31. Wang, B. C. *Meth. Enzym.* **115**, 90-111 (1985).
32. Rice, D. W. *Acta crystallogr.* **A37**, 491-500 (1981).
33. Sussman, J. L., Holbrook, S. R., Church, G. M. & Kim, S.-H. *Acta crystallogr.* **33**, 800-804 (1977).
34. Bricegne, G. *Acta crystallogr.* **A32**, 832-847 (1976).
35. Koller, B. H. & Orr, H. T. *J. Immun.* **134**, 2727-2733 (1985).
36. Marquart, M. & Deisenhofer, J. *Immun. Today* **3**, 160 (1982).
37. Amzel, L. M. & Poljak, R. J. *A. Rev. Biochem.* **48**, 961-997 (1979).
38. Becker, J. W. & Reeke, G. N. *Proc. natn. Acad. Sci. U.S.A.* **82**, 4225-4229 (1985).
39. Vega, M. A., Bragado, R., Ezquerro, A. & Lopez de Castro, J. A. *Biochemistry* **23**, 823-831 (1984).
40. Novotny, J. & Auffray, C. *Nucleic Acid Res.* **12**, 243-255 (1984).
41. McLachlan, A. D. in *Protein Folding* (ed. Jaenicke, R.) 79-99 (Elsevier, North Holland, 1980).
42. Reeke, G. N., Becker, J. W. & Edelman, G. M. *J. biol. Chem.* **250**, 1525-1545 (1975).
43. Krause, K. L., Volz, K. W. & Lipscomb, W. N. *J. molec. Biol.* **193**, 527-553 (1987).
44. Krangel, M. S., Biddison, W. E. & Strominger, J. L. *J. Immun.* **130**, 1856-1862 (1979).
45. Lancet, D., Parham, P. & Strominger, J. L. *Proc. natn. Acad. Sci. U.S.A.* **76**, 3844-3848 (1979).
46. Nathenson, S. G. & Cullen, S. E. *Biochim. biophys. Acta* **344**, 1-25 (1974).
47. Ploegh, H. L., Orr, H. T. & Strominger, J. L. *J. Immun.* **126**, 270-275 (1981).
48. Santos-Aguado, J., Biro, P. A., Fuhrmann, U., Strominger, J. L. & Barbosa, J. A. *Molec. cell. Biol.* **7**, 982-990 (1987).
49. Kimball, E. S. & Coligan, J. E. *Contemp. Top. molec. Immun.* **9**, 1-63 (1983).
50. Terhorst, C., Parham, P., Mann, D. L. & Strominger, J. L. *Proc. natn. Acad. Sci. U.S.A.* **73**, 910-944 (1976).
51. Kaufman, J. F., Auffray, C., Korman, A. J., Shackelford, D. A. & Strominger, J. L. *Cell* **36**, 1-13 (1984).
52. Larhammar, D. *et al. Proc. natn. Acad. Sci. U.S.A.* **79**, 3687-3691 (1982).
53. Sant, A. J., Braunstein, N. S. & Germain, R. N. *Proc. natn. Acad. Sci. U.S.A.* (in the press).
54. Lee, B. & Richards, F. M. *J. molec. Biol.* **55**, 379-400 (1971).
55. Eklund, H. *et al. J. Molec. Biol.* **102**, 27-59 (1976).

The foreign antigen binding site and T cell recognition regions of class I histocompatibility antigens

P. J. Bjorkman^{*†‡}, M. A. Saper^{*}, B. Samraoui^{*}, W. S. Bennett^{*‡}, J. L. Strominger^{*} & D. C. Wiley^{*§}

^{*} Department of Biochemistry and Molecular Biology, Harvard University and Howard Hughes Medical Institute, Harvard University, Cambridge, Massachusetts 02138, USA

[†] Department of Medical Microbiology, Stanford University, Stanford, California 94305, USA

[‡] Present addresses: Department of Medical Microbiology, Stanford University, Stanford, California 94305, USA. (P.J.B.) and Max-Planck-Institute Für Molekulare Genetik, Abteilung Wittmann, D-1000 Berlin 33, FRG (W.S.B.)

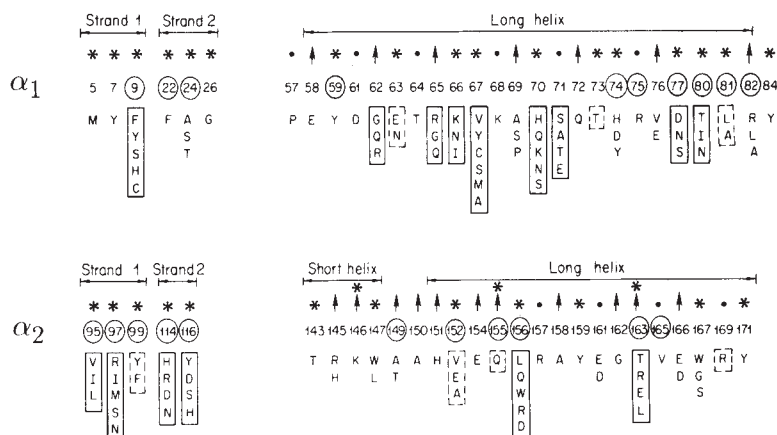
[§] To whom correspondence should be addressed

Most of the polymorphic amino acids of the class I histocompatibility antigen, HLA-A2, are clustered on top of the molecule in a large groove identified as the recognition site for processed foreign antigens. Many residues critical for T-cell recognition of HLA are located in this site, in positions allowing them to serve as ligands to processed antigens. These findings have implications for how the products of the major histocompatibility complex (MHC) recognize foreign antigens.

THE three-dimensional structure of HLA-A2, a human class I histocompatibility antigen, was determined to advance our understanding of how these polymorphic membrane glycoproteins interact with foreign antigens and with receptors on T cells. To trigger lysis of a virus-infected cell, a receptor on a cytotoxic T lymphocyte (CTL) recognizes a complex between a specific viral antigen and a class I histocompatibility molecule¹. A high percentage of CTL are also capable of lysing uninfected cells bearing a foreign (or allo) HLA molecule, a phenomenon known as alloreactivity^{2,3}. The dual recognition of a foreign antigen complexed to HLA (MHC restriction) has remained puzzling, because it seems unlikely that the limited number of

HLA molecules in a given individual could be structurally complementary to the huge number of foreign antigens with which they must complex. Recent evidence, however, has suggested that foreign antigens are not recognized as native three-dimensional structures by T cells, but are processed intracellularly into short peptides⁴ that are presented by histocompatibility molecules. Peptides from foreign antigens have been shown to bind to purified class II histocompatibility molecules^{5,6} at a single site⁷ with micromolar dissociation constants^{5,8}. Because peptides from a viral protein added to uninfected cells will stimulate virus-specific, class I-restricted cell lysis⁹, class I molecules are also expected to bind fragments of viral proteins

Fig. 1 Residues in the vicinity of the recognition site for processed foreign antigen. Listed are accessible residues on the two α -helices and the four central β -strands of the α_1 - α_2 β -pleated sheet that could potentially interact with a bound processed antigen or a T-cell receptor bound to the antigen recognition site (see text and preceding article¹² for details). Amino acids are listed for the HLA-A2 sequence in single-letter code (top line). Amino acids found in other human class I molecules are listed below the HLA-A2 sequence. The locations of these residues on the HLA-A2 structure are noted above the residue numbers, and residues in these positions are shown on the structure of the α_1 and α_2 domains in Fig. 2. *, A residue pointing towards the site; $\uparrow$, residue on an α -helix pointing up (away from the membrane proximal immunoglobulin-like domains); $\bullet$, residue on an α -helix pointing away from the site. Residues enclosed in a rectangular box indicate a position defined as polymorphic^{14,15} in human class I sequences. Residues enclosed in a dashed box indicate a position defined as polymorphic¹⁶ in murine class I sequences. Circled residues are changed in CTL recognition mutants (see Table 1). Residues pointing towards the site are potential ligands for a bound, processed foreign antigen. Some of the ligand residues located on an α -helix might also be accessible to a T-cell receptor. The ligand residues on β -strands would probably be inaccessible to a T-cell receptor when the site is occupied. Residues pointing up on the α -helices can potentially interact with a T-cell receptor bound to the site. Some residues are designated as pointing both into the site and up, indicating that the side chain is in an intermediate position. Residues pointing away from the site could interact with a T-cell receptor. The classification of these residues is based upon their position in the HLA-A2 structure. Amino-acid substitutions in other specificities of HLA could result in a different classification of some of these residues.



to form a complex that is recognized by CTL¹⁰. Some questions which remain are: how does a histocompatibility molecule bind many different peptides? Does a histocompatibility molecule impose a common conformation on peptides to facilitate complex formation and recognition by T-cell receptors? How does the T-cell receptor simultaneously recognize the peptide and the HLA molecule? Why are some peptides presented better than others, leading to a correlation between immunological response to an antigen and certain specificities of histocompatibility molecules¹¹? Does alloreactive recognition of HLA involve peptide binding?

In the preceding article¹², evidence was presented that a prominent groove on the top of HLA-A2 is likely to be the foreign antigen recognition site. Its location, size and shape are consistent with the expectation that it binds a processed (for example, peptide) form of the antigen. Furthermore, in crystals of HLA-A2, the site contains a bound molecule(s) of unknown origin that may be a peptide antigen. In this report we describe the site and the location on the α_1 and α_2 domains of HLA-A2 of the polymorphic amino acids and residues critical for antibody and T-cell recognition of class I histocompatibility molecules. The observation that most polymorphic and T-cell epitope residues are positioned in this site where they could serve as ligands to a processed foreign antigen is further evidence that it serves as the foreign antigen binding site, and it will be referred to as such throughout this article. Some of the other functional residues are located on the edges of the site where they could be recognized directly by a T-cell receptor binding to the antigen recognition site on an HLA molecule.

Antigen recognition site

Figure 1 lists all of the accessible residues on the two helical regions of α_1 and α_2 and on the β -strands located between them that are in the vicinity of the binding site, and therefore likely to be important for T-cell recognition of class I histocompatibility molecules. In Fig. 2, residues of the central β -strands of the α_1 - α_2 β -sheet that form the bottom of the site, and residues on the α -helices facing into the site forming its two sides are coloured red. These residues are potential ligands for antigenic peptides and are designated with an asterisk in Fig. 1. The residues on the top surfaces of the α -helices, pointing up and projecting into solution, are yellow in Fig. 2. These residues are candidates for binding directly to a T-cell receptor (Fig. 1, arrows). Amino-acid substitutions at these locations in other HLA specificities might result in side chains reaching into the

site and contacting an antigenic peptide. Other residues on the helices that are solvent-accessible, but point away from the site, probably do not interact with antigenic peptides but may interact with a T-cell receptor. These are designated only by numbers in Fig. 2 and a dot in Fig. 1.

Polymorphic and conserved residues

Polymorphism at some amino-acid positions of HLA is responsible for the specificity of recognition by T cells, and for the variation in responsiveness to particular foreign antigens. A position is defined as polymorphic if its variability index (number of substitutions at a position divided by a frequency of the most common amino acid at the position) is high¹³. On the HLA-A2 structure, 15 of the 17 positions with a variability index greater than 6 (in 22 human class I sequences^{14,15}) are located in the binding site region for processed antigen (residues 9, 62, 65, 66, 67, 70, 71, 77, 80 in α_1 ; 95, 97, 114, 116, 156, 163 in α_2). Residues 9, 95, 97, 114 and 116 are on the central β -strands of the α_1 - α_2 β -sheet (Fig. 3b; red in Fig. 3a). They point up between the two helical regions, and would be buried if the binding site were occupied, suggesting that the polymorphism of these residues affects interactions with bound antigenic peptides, and not direct interactions with T-cell receptors. Residues 66, 67, 70, 77, 80 and 156 are located on the sides of the helices facing into the site (Fig. 3c; red in Fig. 3a) in positions to be ligands to foreign antigens. Residues 62, 65 and 163 are on the top face of the helices facing solvent (Fig. 3c, yellow in Fig. 3a). They are candidates for making direct contacts with the T-cell receptor. Residue 71 (blue in Fig. 3a) on the α_1 helix does not appear to be in a position to affect the binding of antigenic peptides. The polymorphic residues of the antigenic recognition site are enclosed in rectangles in Fig. 1 (dashed rectangles indicate positions that are polymorphic in murine, but not in human class I molecules). Two polymorphic residues (43, 45; blue in Fig. 3a) are not located on the helices or β -strands that form the binding site, but are found on the outside β -strand of α_1 in an area accessible to solvent. Residue 45 is immediately behind the α_1 α -helix, pointing towards it, and may be able to influence the site. The location of so many polymorphic residues close to or in the site is consistent with the genetic mapping of immune response effects (low response to specific antigens) to histocompatibility molecules¹¹, and with the observation that the binding of peptides from an antigen correlates with the immune response phenotype of the class II molecule for that antigen^{5,6,8}.

Table 1 T-cell epitope residues of human and murine class I molecules

α_1 domain residue	Location on structure	Substitution	Variant (residue changed)	Classification of residue	Ref.
9	β_1^{**} : into site	F $\rightarrow$ Y	CLA (HLA-A2.4) (9)	ligand	20, 30, 43
22	β_2 : into site	Y $\rightarrow$ F	HLA-A2.2Y (9, 43, 95, 156)	ligand	18
23	β_2 : away from site towards β_{2m}	M $\rightarrow$ I	K ^{bm8} (22-24, 30)	silent; see refs 22, 24	18
24	β_2 : into site	E $\rightarrow$ S	K ^{bm8} (22-24, 30)	ligand	18
30	Loop between β_2 and β_3	Y $\rightarrow$ N	K ^{bm8} (22-24, 30)	silent; see refs 22, 24	*
40	Loop between β_3 and β_4	A $\rightarrow$ V	K ^k -TNP restricted	TCR indirect effect?	31
				no affect on allo- or viral-restricted CTL	
43	Loop between β_3 and β_4	E $\rightarrow$ R	M7 (HLA-A2) (43, 95, 156)	silent?; see refs 95, 156	20, 32
59	α -helix: into site	Y $\rightarrow$ H	HLA-B27d (59)	ligand	33
74	α -helix: into site	D $\rightarrow$ Y	HLA-B27F (74, 77, 81)	ligand	34
75	β_2 : into site	R $\rightarrow$ H	K ^{bm8} (75, 77)	TCR	18
77	α -helix: into site	D $\rightarrow$ S	K ^{bm23} (75, 77); K ^{bm11} (77, 80)	ligand	18, 34-37
			K ^{bm3} (77, 89)		
		D $\rightarrow$ N	HLA-B27.2 (77, 80, 81),		
		D $\rightarrow$ S	HLA-B27.3 (77, 152)		
		D $\rightarrow$ N	HLA-B27F (74, 77, 81)		
		D $\rightarrow$ S	HLA-27.4 (77, 114, 116, 152)		
80	α -helix: into site	T $\rightarrow$ N	K ^{bm11} (77, 80);	ligand	18, 35
		T $\rightarrow$ I	HLA-B27.2 (77, 80, 81)		
81	α -helix: into site	L $\rightarrow$ A	HLA-B27.2 (70, 80, 81)	ligand	34, 35
			HLA-B27F (74, 77, 81)		
82	α -helix: points up	L $\rightarrow$ F	R8.313 (H-2K ^b) (82)	TCR	38
89	glycosylated loop	K $\rightarrow$ A	K ^{bm3} (77, 89)	silent?; see ref. 77	18
α_2 domain residues					
95	β_1 : into site	V $\rightarrow$ L	M7(HLA-A2) (43, 95, 156)	ligand	32, 39
		I $\rightarrow$ V	SDM of Aw68 (95, 97)		
97	β_1 : into site	M $\rightarrow$ R	SDM of Aw68 (95, 97)	ligand	39
99	β_1 : into site	Y $\rightarrow$ C	KNE (HLA-A2.4) (99)	ligand	40
107	Loop between β_1 and β_2	G $\rightarrow$ W	SDM of HLA-A3, HLA-Aw68 (107)	TCR, indirect effect?	39, 41
114	β_2 : into site	H $\rightarrow$ D	HLA-B27.4 (77, 114, 116, 152)	ligand	37
116	β_2 : into site	Y $\rightarrow$ F	K ^{bm5,16} (116)	ligand	18, 37, 42
			K ^{bm6,7,8} (116, 121)		
			HLA-B7 (CF)		
		D $\rightarrow$ Y	HLA-B27.4 (77, 114, 116, 152)		
121	Loop between β_2 and β_3	C $\rightarrow$ R	K ^{bm6,7,9} (116, 121)	silent?; see 116	18
149	short α -helix: points up	A $\rightarrow$ T	DK1(HLA-A2) (149, 152, 156)	TCR	20, 32, 43, 44
152	α -helix: into site	E $\rightarrow$ A	K ^{bm1} (152, 155, 156)	ligand/TCR	18, 20, 32, 43
		V $\rightarrow$ E	DK1 (HLA-A2) (149, 152, 156)		36, 37, 44, 45
		V $\rightarrow$ E	HLA-B27.3 (77, 152)		
		E $\rightarrow$ V	E1 (HLA-A3) (152, 156)		
		V $\rightarrow$ E	HLA-B27.4 (77, 114, 116, 152)		
155	α -helix: up and into site	R $\rightarrow$ Y	K ^{bm1} (152, 155, 156)	TCR/ligand	18
156	α -helix: into site	L $\rightarrow$ Y	K ^{bm1} (152, 155, 156)	ligand	18, 20, 32, 43
		L $\rightarrow$ W	DK1 (HLA-A2) (149, 152, 156)		44, 45
		L $\rightarrow$ W	M7(HLA-A2) (43, 95, 156)		
		L $\rightarrow$ Q	E1 (HLA-A3) (152, 156)		
163	α -helix: up and into site	T $\rightarrow$ A	K ^{bm10} (163, 165, 173, 174)	TCR/ligand	18
165	α -helix: points away	V $\rightarrow$ M	K ^{bm10} (163, 165, 173, 174)	TCR; silent? see 163	18
173	α -helix: points down	K $\rightarrow$ E	K ^{bm10} (163, 165, 173, 174)	silent; see 174	18
174	α -helix: points up, not in site	N $\rightarrow$ L	K ^{bm10} (163, 165, 173, 174)	TCR indirect effect?	18

Positions on α_1 and α_2 domains are listed that affect recognition by CTL in a number of Class I molecules. For each position, the location on the HLA-A2 structure (Fig. 6), the amino-acid substitution, the variant name, and the residues changed in that variant are listed. Some of the variant molecules were generated by site-directed mutagenesis (SDM). In the final column, the effect of the residue change is interpreted based on its location on the HLA structure. Residues pointing to the antigen recognition site are classified as 'ligand', and are suggested to be ligands for processed foreign antigens (see residues coloured red in Fig. 6). Residues pointing up on the central β -standards of the α_1 - α_2 β -sheet would be inaccessible if the site were occupied, and are likely to affect T-cell recognition only by the way they alter (or eliminate) peptide binding. Residues on the α -helices pointing in toward the site could be ligands to bound peptides, and in some cases, could be contacted by a T-cell receptor bound to the site. Residues on the α -helices that point up (yellow in Fig. 6), or away from the site (not coloured in Fig. 6) are in positions to be contacted by a T-cell receptor and are classified as 'TCR'. In variants that contain multiple residue changes, we have interpreted certain changes as not contributing to T-cell recognition ('silent', blue in Fig. 6) and directed attention to other residue changes in the same variant. Residues outside the antigen recognition site (blue in Fig. 6) could indirectly affect T-cell receptor binding. The classification of these residues is based upon their position in the HLA-A2 structure. Amino-acid substitutions in other specificities of HLA could result in a different classification of some of these residues.

* Residue 30 substitution (G. Pfaffenbach, *et al.* unpublished data). ** β_N refers to β -sheet strand N.

Fig. 2 Stereo view of the residues in HLA-A2 located in the vicinity of the antigen recognition site. The α -carbon backbone of the α_1 and α_2 domains is shown with side chains of the residues that are close to the site. Side chains pointing in toward the site (red) are potential ligands for bound processed antigens. Side chains on the α -helices pointing up (yellow) or away from the site (not coloured) could interact with a T-cell receptor. Positions are marked with the residue number and amino-acid type in single-letter code (see text and Fig. 1 for details).

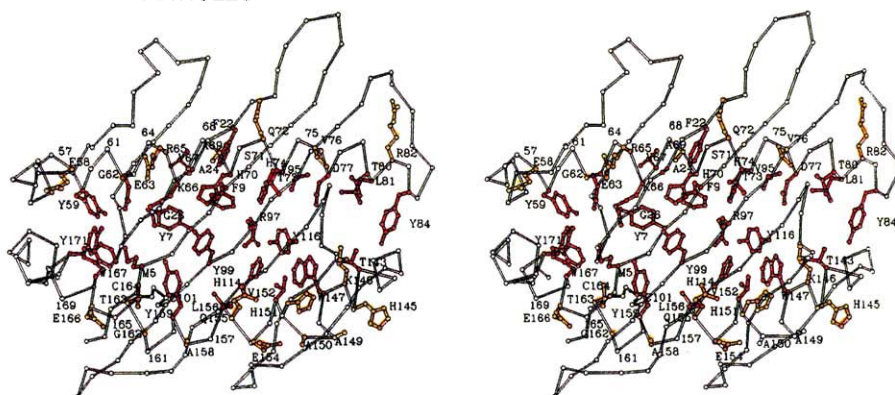


Fig. 3 Location of polymorphic amino acids in the α_1 and α_2 domains of human class I histocompatibility molecules. Polymorphic amino acids are shown highlighted on the α -carbon backbone of HLA-A2 (*a*), or schematically (*b*, *c*). Polymorphic positions are defined as those with a variability index¹³ greater than 6 (using 22 human class I sequences^{14,15}, or 13 murine class I sequences¹⁶). Side chains shown are the amino acids at these positions in HLA-A2. In addition to these polymorphic amino acids, other non-conserved residues (see Fig. 4) could contribute to MHC-restricted T-cell recognition of HLA. *a*, Stereo view of polymorphic amino acids in human class I molecules. Side chains coloured are polymorphic residues in the vicinity of the antigen recognition site (Fig. 2) in positions as potential ligands to antigenic peptides (red; pointing toward site), or to T-cell receptors (yellow; pointing up from site). Residue 163, in an intermediate position, is yellow in this figure (see *c*). Side chains coloured blue are polymorphic residues either outside the antigen recognition site, or not in a position to affect the binding of antigenic peptides (see text and Figs 1, 2 for details). *b*, Schematic view of the polymorphic amino acids located in the α_1 - α_2 β -pleated sheet at the bottom of the antigen recognition site. These residues point up between the α -helices and would be inaccessible if the site were occupied. Their polymorphism presumably affects interactions with bound processed foreign antigens. The β -strands are represented by an arrow in the amino to carboxy direction. Polymorphic positions are indicated by a solid circle (human) or dashed circle (murine). Polymorphic residues 114 and 116 are separated by residue 115, a conserved Gln, that points down to interact with Trp 60 in the membrane-proximal, β_2 -microglobulin domain. *c*, Schematic view of the polymorphic amino acids located in the antigen recognition site on the α_1 and α_2 domain α -helices. The helices are represented as rectangles with arrows indicating the amino to carboxy direction. Polymorphic residues pointing in, towards the antigen recognition site, are represented as circles on the right (α_1) or left (α_2) side of the rectangles. Polymorphic residues which point up into solution are depicted as circles in the centre of the rectangle. Residues 155 and 163 are depicted as pointing both up and towards the site. Residue 71 (square) points down towards the α_1 - α_2 β -sheet and appears inaccessible. Polymorphic positions are indicated by a solid circle (human) or dashed circle (murine).

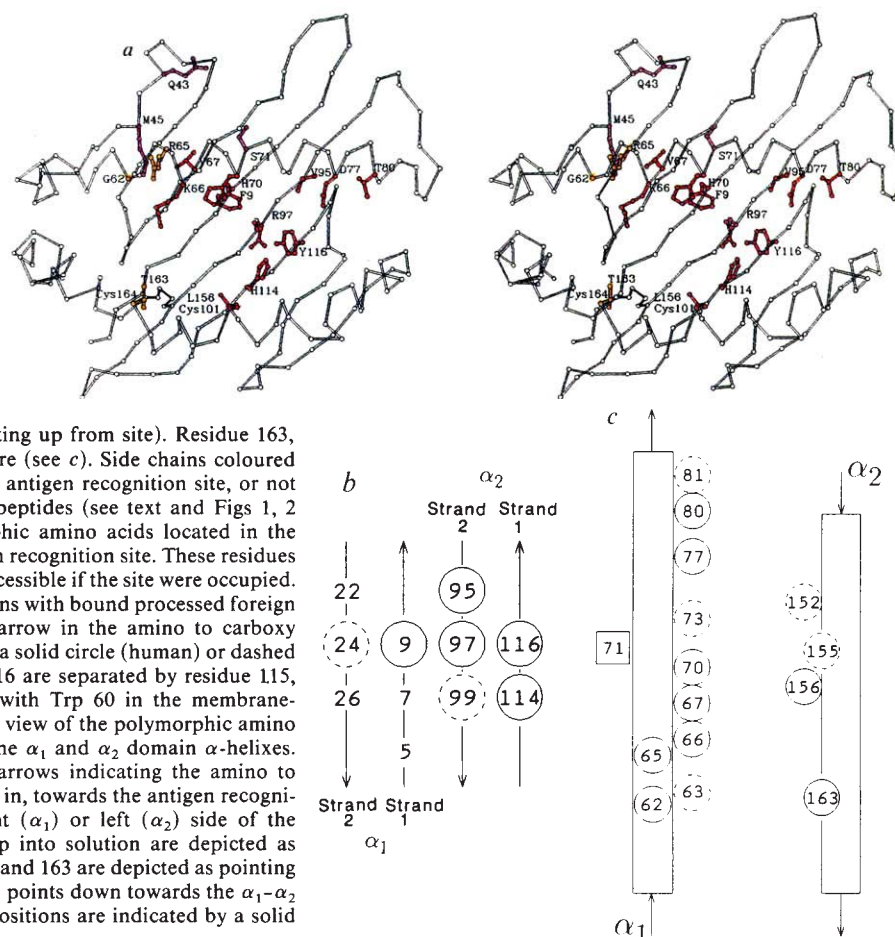
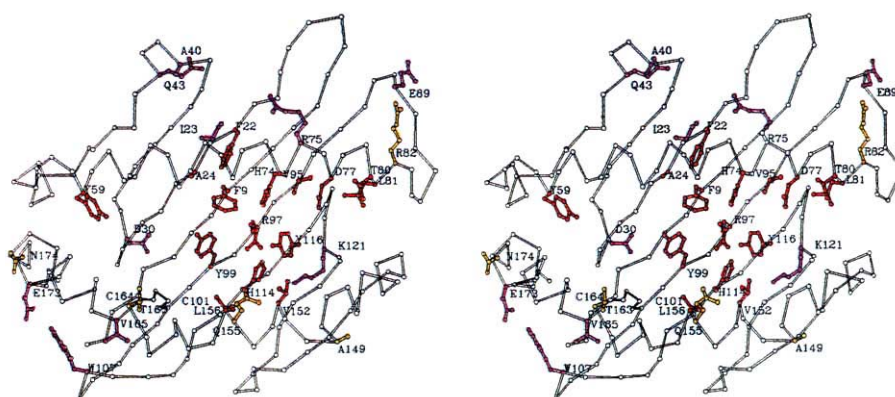


Fig. 6 Stereo view of the T-cell epitopes of human and murine class I molecules. The α -carbon backbone of the α_1 and α_2 domains is shown with side chains of those residue demonstrated to be critical for recognition by alloreactive and MHC-restricted T cells. Side chains shown are the amino acids in these positions in HLA-A2. Residues pointing toward the antigen-recognition site (red) are potential ligands for processed foreign antigens. Residues on the α -helices that point up into solution (yellow) or away from the site (not coloured) could interact with a T-cell receptor bound to the site. Other residues implicated in T cell recognition are coloured blue. Residues 155 and 163 are in intermediate positions and could have been coloured either red or yellow. (See text and Table 1 for an interpretation of the effect of residue changes on CTL recognition of class I molecules.)



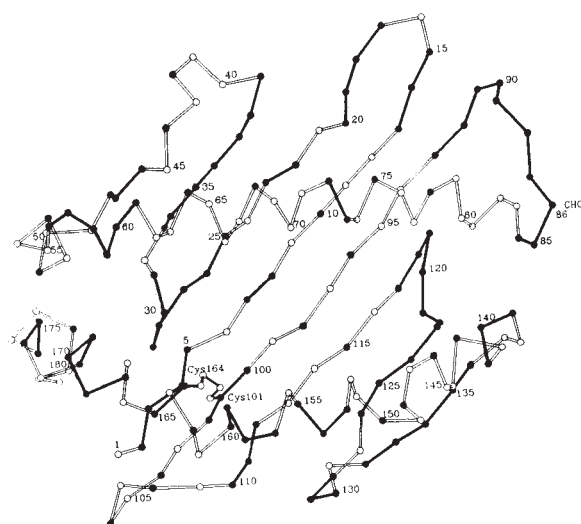


Fig. 4 Location of conserved amino acids in the α_1 and α_2 domains of human class I histocompatibility molecules (sequences from refs 14, 15). Invariant residues are shown as solid circles on the α -carbon backbone of HLA-A2. Open circles indicate positions of non-conserved (but not necessarily polymorphic, see Fig. 3) amino acids. The conserved tetrapeptide RFDS (residues 35–38) in the α_1 domain has been suggested to interact with CD8, a T-cell accessory molecule⁴⁶ based on the similarity between this sequence and the tetrapeptide involved in the cell attachment site of fibronectin⁴⁷. The RFDS sequence is located on β -strand 3 of α_1 . R35 is involved in a salt bridge to D53 of the membrane-proximal β_2 -microglobulin domain. F36 is buried underneath the α_1 helix. This sequence does not appear accessible for interactions with other molecules, although conformational changes upon binding T cells cannot be ruled out. An SDGR sequence also suggested to be important for molecular interactions⁴⁶ is located on an exposed loop at 105–108, but is not conserved (SDWR on HLA-A2).

Conserved residues

Ten residues pointing into the antigenic recognition site are completely conserved in 22 human sequences^{14,15} (M5, Y7, F22, G26, Y59, Y84, T143, K146, Y159, Y171). In 13 murine sequences, they are also conserved or have one occurrence of another amino acid¹⁶. These residues, which include five tyrosines, may recognize a constant feature of processed antigens. A pattern of every third or fourth residue is conserved on the long helix in the α_1 domain of HLA (57, 58, 61, 64, 72, 73, 75, 78, 84). These residues (excluding 78 and 84) form a completely conserved face on this helix, pointing up into solution or away from the antigen-recognition site. A similar patch (146, 150, 154, 155, 158, 162, 165, 169) is conserved on the long helix of α_2 . Whether these conserved amino acids are recognized directly by constant features on T cells or are conserved to maintain the HLA structure cannot yet be determined.

The locations of invariant residues on the α_1 and α_2 domains of HLA-A2 are shown in Fig. 4, where it is evident that several loops on the molecule are highly conserved. Most of the non-conserved residues are located in and around the binding site, suggesting that the location of most of the variability in HLA molecules has been selected to generate an ability to present many different peptides.

Antibody binding sites on HLA

The residues comprising some serological epitopes were determined by comparing the pattern of substitutions in HLA subtype sequences with the distribution of a particular epitope (reviewed in ref. 17). Residues that were identified as critical for the binding of monoclonal antibodies to human or murine class I molecules are shown in Fig. 5 (see legend). All of these residues are located

on the α -helices in the α_1 and α_2 domains, with the exception of E89 and W107, which are found on exposed loops. In each epitope at least one residue is solvent-exposed (62, 75, 80, 82, 89, 107, 149, 152, 161, 177) and appears accessible to direct recognition by an antibody. But one cannot exclude the possibility that some substitutions are not recognized directly, and instead cause conformational changes in other residues which affect antibody binding. None of the polymorphic residues located at the bottom of the antigen recognition site have been demonstrated to be critical for antibody recognition of class I molecules. But some of these residues, which would be buried if the site were occupied, are involved in T-cell recognition of HLA (see below). Figure 5b shows residues identified to be involved in antibody binding to β_2 -microglobulin.

CTL mutants and variants

Specific residues involved in the recognition of murine class I molecules by cytotoxic T lymphocytes (CTL) have been identified by characterizing spontaneous mutations in the *H-2K^b* (class I) gene from mutants selected for skin graft incompatibility with individuals of their parental haplotype (reviewed in ref. 18). In human class I proteins, residues have been identified by isolating subtype variants from serologically-defined HLA specificities on the basis of differential reactivity with CTL (ref. 19; see ref. 20 for review). Most of the human and murine variant molecules have two or three amino-acid substitutions, and in some cases, these changes have resulted from a gene conversion event involving another class I gene¹⁸. Table 1 lists the residues in murine and human class I molecules that are changed in CTL variants and mutants, their location on the HLA-A2 structure, and an interpretation of the probable consequences of the substitution.

The locations of residue changes in CTL variants fall into three categories: (1) residues in the antigenic recognition site defined as potential ligands to an antigenic peptide (red in Fig. 6: 9, 22, 24, 59, 74, 77, 80, 95, 97, 99, 114, 116, 152, 156). Within this category, those residues located at the bottom of the antigenic recognition site (9, 22, 24, 95, 97, 99, 116) would be expected to be covered by a bound antigen, and are therefore likely to be distinguished by CTL only by the way they alter (or eliminate) peptide binding. (2) Residues in the antigenic recognition site that point up or away from the site (75, 82, 149, 155, 163, 165; yellow in Fig. 6). Substitutions at these positions could affect interactions with a T-cell receptor bound to the antigen recognition site. (3) Residues outside the antigen recognition site (blue in Fig. 6: 40, 43, 89, 107, 121, 173, 174). Residues 107, 173 and 174 are near one end of the site where they may affect T-cell receptor binding. The other residues are found on loops. With the exception of residue 40, these mutations (43, 121, 89) are found in variants that also have at least one other change in a residue at a potential ligand position in the antigen recognition site. Therefore they may not be residues critical for T-cell recognition (see Table 1).

Alloreactive and MHC-restricted recognition

The murine CTL variants were selected on the basis of alterations in recognition by alloreactive CTL. Some have also been tested in virus-specific, MHC-restricted killing assays¹⁸. In almost all cases, at least one MHC-restricted T-cell line was found whose reactivity was affected by the mutation. The residue changes reported to have the greatest effect on class I-restricted CTL are located at positions 152, 155 and 156, and 77–80 (refs 18, 20). These regions include residues that are potential ligands in the binding site.

The recognition of a foreign HLA by alloreactive T cells is thought to be a cross-reaction occurring because the structure of a self MHC molecule plus antigen (self + x) resembles a non-self (allo) MHC molecule²¹. A T-cell receptor could directly contact the polymorphic residues on an allo HLA molecule to distinguish it from self, or the receptor could recognize the

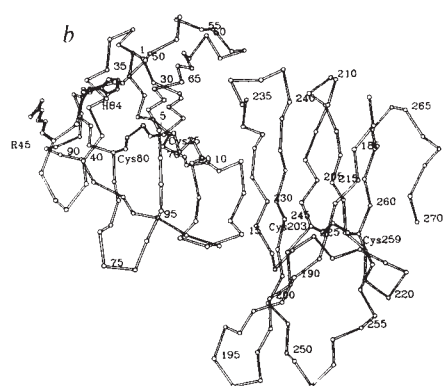
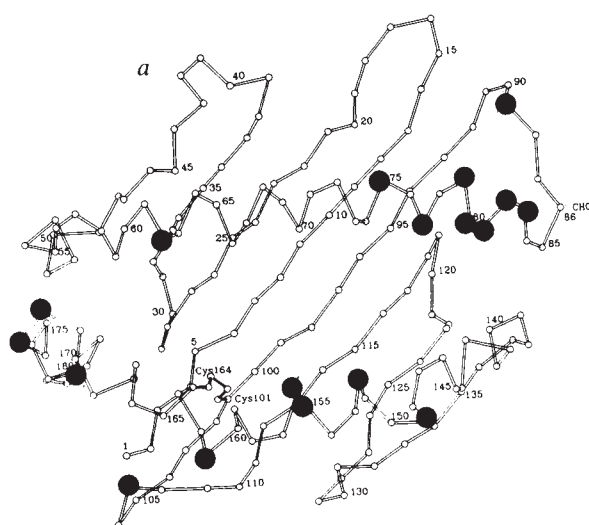


Fig. 5 Serological epitopes. **a**, Serological epitopes on the α_1 and α_2 domains of human and murine class I molecules. Residues demonstrated to be critical in the binding of monoclonal antibodies are shown with solid circles on α_1 and α_2 . The epitopes involve residues: (i) 62. Antibody: MA2.1 (HLA-A2, -B17-specific)⁴⁸ correlates with A3, A11 crossreactivity¹⁴. (ii) 77. Antibody: B27M2 (HLA-B27-specific)³⁶. (iii) 79, 89. H-2K^b specific antibodies⁴⁹. (iv) 79-83. Correlates with HLA-BW4, -BW6 epitopes⁵⁰. (v) 82. Antibody: EH-144 (H-2K^b-specific)³⁸. (vi) 107. Antibody: PA2.1 (HLA-A2, Aw69-specific)^{39,41}. (vii) (142, 145) 149, 152, 156. Antibody: CR11-351 (HLA-A2, A28-specific). Epitope proposed to be caused by the effect of residues 142 and 145 on the 149-156 region⁵³. (viii) 152, 155, 156. H-2K^b-specific antibodies⁴⁹. (ix) 161. Antibody: BB7.1, PA2.1 (HLA-A2, Aw69-specific). (Residues 107 and 161, which are both critical for PA2.1 binding to HLA-A2 (ref. 42) are separated by 14 Å.) (x) 177, 178, 180. Antibody: MB40.2 (HLA-B40, B7-specific)⁵¹. Site-directed mutagenesis of HLA-A2⁵⁴ has also located epitopes for PA2.1, MA2.2 and BB7.2 (HLA-A2 specific) at residue 107, MA2.1 (HLA-A2, B17 specific) at residues 62-66, SFR-8-B6 (Bw6 specific) at residue 79 and MB40.2 and MB40.3 (HLA-B7, B40 specific) at residues 177-180 (ref. 55). **b**, Epitopes located on β_2 -microglobulin (β_2 m): a stereo drawing of the α -carbon chain of the α_3 and β_2 m domains showing β_2 m residues (solid bonds) critical for antibody binding. (i) 45. Antibody: BBM1 (reacts with HLA-associated and free β_2 m⁵²). (ii) 84. Antibody: PAC.M1 (reacts with an epitope common to β_2 m immunoglobulin light chains, class I heavy chains, class II α and β chains, and papain solubilized HLA⁵³).

molecule complexed with a peptide. Although either mechanism is possible, the observation that the majority of alloreactive CTL variant residues are in the peptide-binding site, rather than on surfaces facing towards solution, suggests that alloreactivity may involve the recognition of a complex between the allo HLA and some peptide. Such a peptide could be derived from the foreign HLA itself or another molecule. Although it is conceivable that substituted residues on an allo HLA molecule are recognized on a peptide derived from itself, none of these residues (in the studies cited above) are buried in the structure of HLA-A2. Thus, it is not necessary to fragment the allo molecule to expose these residues for CTL recognition.

Recognition of foreign antigens

Several observations suggest that the prominent groove between the α_1 and α_2 domain α -helices is a foreign antigen recognition site. The site is located on a surface distal from the membrane end of the HLA molecule, in a position for recognition by receptors from the surface of another cell. The site, ~25 Å long by 10 Å wide by 11 Å deep, has a size and shape consistent with the expectation that, by analogy with class II molecules, class I molecules bind a processed antigen in the form of a peptide. Electron density representing an unknown molecule, possibly a bound peptide antigen, is found in the site in two crystal forms of HLA-A2 (see preceding article¹²). Many of the highly polymorphic amino acids that are responsible for recognition by T cells and haplotype-specific associations with antigens are located in the site. A majority of the amino-acid substitutions in murine and human class I molecules that alter recognition by T cells face into the site.

Peptide binding

Synthetic peptides bind to purified murine⁵⁻⁸ and human (T. Jardetzky *et al.*, unpublished data) class II molecules, presumably mimicking processed foreign antigens. Peptide binding to class I molecules has not been reported, but synthetic peptides serve as antigens in virus-specific, class I-restricted CTL recognition⁹. Because class I and class II molecules have homologous structures²² and T cells specific for either class of molecule use the same receptors^{23,24}, the type of interaction described between peptides and class II molecules is assumed to apply to peptides and class I molecules¹⁰. In an individual, either class of histocompatibility molecule must bind a very large number of foreign peptides, possibly in a common conformation. An α -helical conformation has been suggested for bound peptides based on the effect of amino acid substitutions on CTL recognition^{25,26}. An amphipathic α -helix was suggested by analysing the sequences of CTL epitopes^{27,28}. In both cases, one face of a peptide α -helix was envisioned to contact the class II molecule, leaving the other face to be contacted by a T-cell receptor.

The shape of the antigen recognition site on HLA-A2 suggests that a bound peptide would make contacts on three of its surfaces, to the bottom and two sides of the groove, rather than to one surface. The site contains a number of tyrosine and many polar or charged residues, as well as non-polar amino acids (Figs 1, 2), obviating a need for strict amphipathicity of a bound peptide. The site would accommodate an α -helical peptide of ~20 residues, or an extended polypeptide chain of ~eight residues. If the actual conformation of the bound peptide were bent, or only partly helical with unfolded ends, it could have an intermediate number of residues. One end of the groove

(near amino acid 77) appears open and may allow longer antigens to extend out of the site in that direction. The other end contains four conserved residues (M5, Y5, Y59, Y159) with three tyrosine hydroxyl groups directed into the site. These conserved potential ligand residues could recognize a common feature of processed antigens, such as the main chain atoms of a peptide, a common class of terminal residue, or a chemical adduct associated with antigen processing²⁹.

In a recent report, residues that contact the T cell and residues that contact a class II molecule were defined on the lysozyme 52–61 peptide²⁶. When the peptide was modelled as an α -helix and energy-minimized, a conformation was found in which residues important for T-cell contact oriented in one direction, and residues important for contact to the class II oriented in the opposite direction. This segregation of residues suggested that the lysozyme 52–61 peptide may bind to a class II molecule as an α -helix with the T-cell contact residues facing solvent²⁶. We suggest that as two of the class II contact residues on the peptide (residues 52 and 61) are 10 residues (2.5 turns) apart, they would be on opposite sides of a standard α -helix. If class II molecules have an antigen binding groove like that on HLA-A2 (see preceding article¹²), then lysozyme residues 52 and 61 might contact opposite sides of that groove; residue

52 to the α -helix on one domain and residues 61 to the α -helix on the other domain (J. Brown *et al.*, manuscript in preparation).

The molecule or mixture of molecules bound to crystalline HLA-A2 might be processed peptides, but their composition and conformation cannot be unambiguously determined¹². However, finding molecules bound to the antigen recognition site in crystals of HLA-A2 implies that they remained bound to HLA during the entire purification and crystallization procedure; an observation consistent with the slow kinetic off-rates (of the order of days) reported for peptides bound to class II molecules^{2,6} and may indicate that a conformational change is required to release bound peptide. This may suggest that a peptide is always bound to histocompatibility antigens, and that its displacement is an essential feature of cellular immunity.

We thank Tom Garrett for help with the figures and our colleagues at Harvard and Stanford for helpful discussions. P.J.B. thanks Mark Davis for support and laboratory space during the year 1986–1987. B.S. thanks the Algerian Ministère de l'Enseignement et de la Recherche Scientifique for a postdoctoral leave. P.J.B. held an American Cancer Society postdoctoral fellowship during part of the work. The research was supported by the NIH and the Howard Hughes Medical Institute.

Received 17 August; accepted 27 August, 1987.

- Zinkernagel, R. M. & Doherty, P. C. *Nature* **248**, 701–702 (1974).
- Klein, J. in *Biology of the Mouse Histocompatibility Complex* (Springer, Berlin, 1975).
- Bach, F. H. & van Rood, J. J. *New Engl. J. Med.* **295**, 806–813 (1976).
- Shimonkevitz, R., Kappler, J., Marack, P. & Grey, H. *J. exp. Med.* **158**, 303–316 (1983).
- Babbitt, B., Allen, G., Matsueda, E., Haber, E. & Unanue, E. *Nature* **317**, 359–361 (1985).
- Buus, S. *et al. Proc. natn. Acad. Sci. U.S.A.* **83**, 3968–3971 (1986).
- Guillet, J. G., Lai, M.-Z., Briner, T. J., Smith, J. A. & Geffter, M. L. *Nature* **324**, 260–263 (1986).
- Buus, S., Sette, A., Colon, S. M., Jenis, D. M. & Grey, H. M. *Cell* **47**, 1071–1077 (1986).
- Townsend, A. R. M. *et al. Cell* **44**, 959–968 (1986).
- Forman, T. *Adv. Immun.* (in the press).
- McDevitt, H. O. & Tyan, M. L. *J. exp. Med.* **128**, 1–11 (1968).
- Bjorkman, P. J. *et al. Nature* **329**, 506–512 (1987).
- Wu, T. T. & Kabat, E. *J. exp. Med.* **132**, 211–250 (1970).
- Cowan, E. P., Jelachich, M. L., Biddison, W. E. & Coligan, J. E. *Immunogenetics* **25**, 241–250 (1987).
- Gussow, D. *et al. Immunogenetics* **25**, 313–322 (1987).
- Immun. Today*, Centre page diagram, 13 (Elsevier, Cambridge, U.K.).
- Parham, P. *Immun. Res.* **6**, 153–178 (1987).
- Nathenson, S. G., Geliebter, J., Pfaffenbach, G. M. & Zeff, R. A. *Rev. Immun.* **4**, 471–502 (1986).
- Biddison, W. E., Ward, F. E., Shearer, G. M. & Shaw, S. J. *Immun.* **124**, 548–552 (1980).
- Lopez de Castro, J., Barbosa, J. A., Krangel, M. S., Biro, A. & Strominger, J. L. *Immunol. Rev.* **85**, 149–168 (1985).
- Bevan, M. J. *Proc. natn. Acad. Sci. U.S.A.* **74**, 2094–2098 (1977).
- Kaufman, J. F., Auffray, C., Korman, A. J., Shackelford, D. A. & Strominger, J. L. *Cell* **36**, 1–13 (1984).
- Rupp, R., Acha-Orbea, H., Hengartner, H., Zinkernagel, R. & Joho, R. *Nature* **315**, 425–427 (1985).
- Marrack, P. & Kappler, F. *Adv. Immun.* **38**, 1–30 (1986).
- Watts, T. H., Garipey, J., Schoolnik, G. K. & McConnell, H. M. *Proc. natn. Acad. Sci. U.S.A.* **82**, 5480–5484 (1985).
- Allen, P. M. *et al. Nature* **327**, 713–715 (1987).
- Delisi, C. & Berzofsky, J. A. *Proc. natn. Acad. Sci. U.S.A.* **82**, 7048–7052 (1985).
- Berkower, I., Buckenmeyer, G. K. & Berzofsky, J. A. *J. Immun.* **136**, 2498–2503 (1986).
- Nairn, R., Spengler, M. L., Hoffman, M. D., Solvay, M. J. & Thomas, D. W. *J. Immun.* **133**, 3225–3234 (1984).
- Ezquerro, A. *et al. J. Immun.* **137**, 1642–1649 (1986).
- Karmann, G. *et al. Immunogenetics* **22**, 35–38 (1985).
- Mattson, D. H. *et al. Immunogenetics* (in the press).
- Rojo, S., Aparicio, P., Hansen, J., Choo, S. Y. & Lopez de Castro, J. A. *J. Immun.* (in the press).
- Rojo, S., Aparicio, P., Choo, S. Y., Hansen, J. A. & Lopez de Castro, J. A. *J. Immun.* **139**, 831–836 (1987).
- Vega, M. A. *et al. Proc. natn. Acad. Sci. U.S.A.* **82**, 7394–7398 (1985).
- Vega, M. A. *et al. J. Immun.* **135**, 3323–3332 (1985).
- Vega, M. A., Bragado, R., Ivanyi, P., Palaez, J. L. & Lopez de Castro, J. A. *J. Immun.* **137**, 3557–3565 (1986).
- Nakagawa, M., Zeff, R. A., Geier, S. S., Bluestone, J. A. & Nathenson, S. G. *Immunogenetics* **24**, 381–385 (1986).
- Salter, R. D., Clayberger, C., Lomen, C. E., Krensky, A. M. & Parham, P. *J. exp. Med.* **166**, 283–288 (1987).
- Domenech, N., Ezquerro, A., Castano, R. & Lopez de Castro, J. A. *J. Immun.* (submitted).
- Layet, C. *et al. J. Immun.* **138**, 2197–2201 (1987).
- Taketani, S., Krangel, M. S., Pious, D. & Strominger, J. L. *J. Immun.* **131**, 2935–2938 (1983).
- Holmes, N., Ennis, P., Wan, A. M., Denney, D. W. & Parham, P. *J. Immun.* **139**, 936–941 (1987).
- Krangel, M. S., Biddison, W. E. & Strominger, J. L. *J. Immun.* **130**, 1856–1862 (1983).
- Cowan, E. P., Jordan, B. R. & Coligan, J. E. *J. Immun.* **135**, 2835–2841 (1985).
- Auffray, C. & Novotny, J. *hum. Immun.* **15**, 381–390 (1986).
- Yamada, K. M. & Kennedy, D. W. *J. cell. Biochem.* **28**, 99–104 (1985).
- Ways, J. P. & Parham, P. *Biochem. J.* **216**, 423–432 (1983).
- Hammerling, G. J., Rusch, E., Tada, N., Kimura, S. & Hammerling, U. *Proc. natn. Acad. Sci. U.S.A.* **79**, 4737–4741 (1982).
- Wan, A. M., Ennis, P., Parham, P. & Holmes, N. J. *Immun.* **137**, 3671–3674 (1986).
- Ways, J. P., Lawlor, D. A., Wan, A. M. & Parham, P. *Immunogenetics* **25**, 323–328 (1987).
- Parham, P., Androlewicz, M. J., Holmes, N. J. & Rothenberg, B. E. *J. biol. Chem.* **258**, 6179–6186 (1983).
- Parham, P., Lutz, P. M. & Cresswell, P. *Immunogenetics* **25**, (in the press).
- Santos-Aguado, J., Biro, P. A., Fuhrmann, U., Strominger, J. L. & Barbosa, J. A. *Molec. Cell Biol.* **7**, 982–990 (1987).
- Santos-Aguado, J., Barbosa, J. A., Mentzer, S., Burakoff, S., Biro, P. A. & Strominger, J. L. *Abstracts, Fifth HLA and H-2 Cloning Workshop*, Geneva, Switzerland (1986) and manuscript in preparation.
- Maryanski, J. L., Pala, P., Corradin, G., Jordon, B. R. & Cerottini, J.-C. *Nature* **324**, 578–579 (1986).

An ancient planetary nebula surrounding the old nova GK Persei

M. F. Bode*, E. R. Seaquist†, D. A. Frail†,
J. A. Roberts*, D. C. B. Whittet*, A. Evans‡
& J. S. Albinson‡

* School of Physics and Astronomy, Lancashire Polytechnic,
Corporation Street, Preston PR1 2TQ, UK

† Department of Astronomy, University of Toronto,
60 St George Street, Toronto, ON M5S 1A7 Canada

‡ Department of Physics, University of Keele, Staffordshire,
ST5 5BG, UK

Classical novae are a subset of the cataclysmic variable (CV) class of objects which undergo outbursts with peak luminosities $\sim 5 \times 10^{37}$ – 5×10^{38} erg s⁻¹ every 10^4 – 10^5 years. Around 10^{-5} – $10^{-4} M_{\odot}$ of material is ejected at velocities typically 1,000 km s⁻¹ at outburst. The central system is a semi-detached binary containing a white dwarf and (usually) a late-type, main-sequence companion. The companion fills its Roche lobe, and loses material, via an accretion disk, onto the surface of the white dwarf. Orbital periods are normally found to be in the range 3–10 h with CV orbits generally of low eccentricity (see ref. 1 for a review). As part of a continuing programme of observations of the old nova GK Persei, we examined Infrared Astronomical Satellite (IRAS) images in the region of the nova, and discovered extended emission in the far infrared. These observations are interpreted in terms of an ancient planetary nebula ejected from the central binary system. If this interpretation is correct, then several unique phenomena associated with GK Per may be explained. In addition, GK Per would then provide valuable clues to the evolutionary status of classical novae, and the later stages of planetary nebula formation.

GK Persei underwent a classical nova outburst in 1901, reaching visual magnitude (*V*) of 0.2 on 22 February 1901. Subsequently, it declined at 0.23 mag per day, reaching a minimum typically of *V* = 13.1 (ref. 2). In contrast with other objects of this type, GK Per shows dwarf nova-like outbursts of ~ 2 mag, the latest of which was in November 1986 (ref. 3). It is also unusual in that it has by far the longest orbital period among the classical novae (1.904 days, ref. 4; 1.997 days, ref. 5) and an evolved secondary with spectral type K2IV (refs 4, 5). There is controversy about the eccentricity of the orbit⁵, but several authors have suggested it is highly eccentric ($e = 0.39$, ref. 4). The primary is a white dwarf, with mass estimated by Watson *et al.*⁶ to be $>0.72 M_{\odot}$, by Crampton *et al.* to be $0.9 M_{\odot}$ and by Bianchini *et al.*⁴, as $1.3 M_{\odot}$. The orbital inclination is at present uncertain, but it is suggested that it is high⁷. Bianchini and Sabbadin⁸ suggest a value of $i = 66^{\circ}$.

A month after the 1901 outburst, expanding nebulosities were observed in the vicinity of the nova⁹. The apparent superlight expansion of these nebulosities was explained by Couderc¹⁰ in terms of reflection from dust grains lying in a plane crossing the line of sight to the nova. Only for one other object have such phenomena been claimed to have been seen¹¹, despite modern searches for the effect around recent novae¹². Radio observations of GK Per in 1983 (ref. 13) showed that the remnant associated with the ejecta is a non-thermal radio source, apparently unique among classical novae¹⁴. These observations suggested that we are witnessing the interaction of the ejecta with a dense ambient medium, possibly in the form of an interstellar cloud¹⁴.

A campaign of observations was begun in 1984 to explore further the radio emission from the remnant, and to determine the nature of the ambient medium with which the ejecta from the 1901 outburst may be interacting. These results will be reported in a forthcoming paper (E.R.S. *et al.*, in preparation). As part of this investigation, IRAS Skyflux images¹⁵ of the region

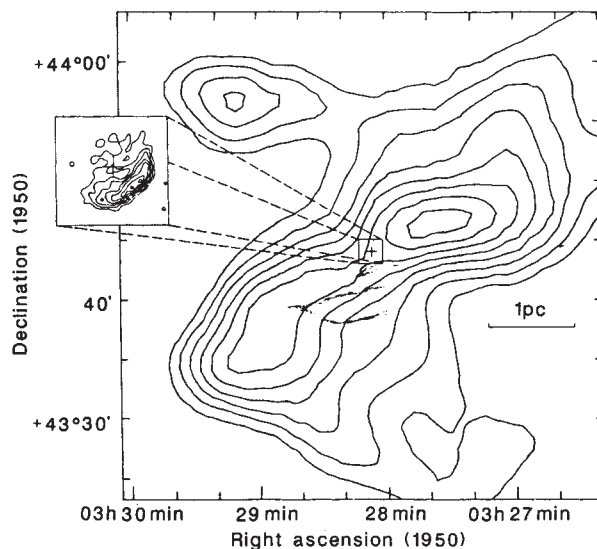


Fig. 1 IRAS 100- μ m map of the region around GK Per. The position of the nova is marked. Contours range from 1.5 MJy sr⁻¹ (3σ level) to 5 MJy sr⁻¹ in intervals of 0.5 MJy sr⁻¹. The peak intensity of 5.26 MJy sr⁻¹ is marked. Superimposed is a sketch of the disposition of reflection nebulosity from a Lick Observatory plate taken on 12–13 November 1901. The inset shows a 5 GHz radio map of the non-thermal radio emission of the central interaction region as observed in August 1984.

around the nova were examined. Extended emission was detected in both the 60- and 100- μ m bands centred around the nova. Subsequently, the database containing raw images from the satellite was consulted. It transpired that GK Per lay in a region of sky that had been part of an unrelated deep survey, and no less than 25 scans across the region had been made. This enabled us to search $\sim$ two to three times deeper in surface brightness than could be achieved with Skyflux images. It also meant that spatial resolution was increased from around 5' in Skyflux images to $\sim 2'$ and $1'$ in the co-added raw data at 100 and 60 μ m respectively. Inspection of the resulting 12 and 25- μ m images, which were degraded to the resolution of the 60- μ m map, suggests that the source was detected here as well, although the highest contours of surface brightness were only approximately 6σ above the noise level. The total flux at 100 μ m down to the 3σ surface brightness contour in the map was determined to be 217.1 Jy, after subtraction of a mean background. The major uncertainty in this figure is the calibration error of the IRAS detectors, which amounts to $\sim 20\%$ in this band, 15% at 60 μ m, 10% at 25 μ m and 5% at 12 μ m (D. H. Hughes, personal communication). Subtracting other values of background, which were mean or sloping fits to whole scans, did not yield results that were different by more than the systematic error.

Figure 1 shows the 100- μ m map of the region, with the disposition of the reflection nebulosity sketched from a photograph taken on 12–13 November 1901 with the Crossley reflector at Lick Observatory, and the scale and orientation of the non-thermal radio emission given in the inset. The 60- μ m map has a similar appearance, but with lower values of surface brightness. The most striking features of these maps are the bipolar nature of the far-infrared emission, and its almost symmetrical displacement about GK Per, with the nova situated on a saddle point between the peaks of emission. The source to the north-east of the nova is interesting in that it has similar colour temperature to the main emission, and it lies in a roughly orthogonal direction to the axis of the IRAS nebulosity. But evidence for its association with GK Per is otherwise lacking and we will not consider it further at this stage.

The ratio of fluxes at 60 and 100 μ m ($f_{60}/f_{100} = 0.175 \pm 0.015$) is similar to that found for the infrared cirrus that is associated

with thermal emission from grains with radii 0.01–0.1 μm in the interstellar medium, heated by the ambient stellar radiation field¹⁶. Assuming that the emissivity of the grains varies as $\sim \nu^2$ (ref. 17) for a graphite or silicate grain in this wavelength range, a grain temperature of $T = 22 \pm 1$ K is derived. The dominant uncertainty here is again the absolute calibration of the IRAS bands. Using this temperature, we would not have expected to have detected the source at 12 and 25 μm at the level of ~ 10 Jy that appears to have been found, but the infrared cirrus also shows an unexpected excess in these bands, attributed to the presence of an additional population of extremely small grains with radii 3–10 Å (ref. 18). The total flux from the nebula at 100 μm can then be used to derive the mass of dust in the cloud from

$$M_d = 4\pi a d^2 \rho / 3 B(T_g) Q_{\text{abs}}$$

where a is grain radius, d is the distance to the source, ρ is the bulk density of the grain material, and Q_{abs} is the grain absorption efficiency, given for graphite spheres in this size range at 100 μm by

$$Q_{\text{abs}} = 1.9 \times 10^{-4} (a/0.01 \mu\text{m}) \text{ (ref. 17)}$$

With $d = 470$ pc (ref. 19), $\rho = 2.2 \text{ g cm}^{-3}$, and $T_g = 22$ K, we find $M = 0.058 M_\odot$ independent of grain size. For silicate grains the calculated mass is $0.078 M_\odot$. For comparison, we have also made 21-cm H I observations, reported separately (E.R.S. *et al.*, in preparation). The H I mass is estimated to be $\geq 0.6 M_\odot$ leading to a dust to gas ratio of ≤ 0.09 , assuming solar composition. This limiting value is an order of magnitude higher than the average value for the interstellar medium in the solar neighbourhood²⁰, suggesting that the abundance of dust may be enhanced in the nebula associated with GK Persei. This ratio should, however, be approached with some caution as the H I mass is a lower limit imposed by the fact that more extended emission may have been resolved out in our VLA maps, and the dust mass is derived in an entirely different way from the techniques used by Savage and Mathis²⁰.

Measurements of f_{60}/f_{100} across the nebula show no significant changes in temperature. This, together with the bipolar appearance of the emission, and the fact that the nova is situated on a saddle point, suggests we are in fact viewing an almost edge-on torus of material surrounding GK Per. The origin of this toroid must then be explored.

First, we consider the possibility that the nova lay near an interstellar cloud before previous outbursts and that the appearance in the IRAS maps is due to sweeping up of grains, or grain destruction following shock passage after a nova event. The symmetrical displacement of the peaks in the latter case would then be due to the resulting cloud being heated by the radiation flux of the quiescent nova system.

As the ejecta from the nova are expelled from the system at far greater velocity than the sound speed in the ambient medium, a shock system will be set up. Dust grains will then be subject to either (1) sputtering or (2) sweeping up in the shock. Once a shock has swept up several times its own mass of material, the ejecta can be assumed to take no further part in the dynamics, and the path of the blast wave can be described using the well-known Sedov solution for an adiabatic shock moving into a uniform medium, namely

$$r_s = \xi (\epsilon_0 / \rho_0)^{1/5} t^{2/5} \text{ (ref. 21)}$$

Here r_s is the shock radius at time t , ϵ_0 is the input energy ($\sim 10^{45}$ erg for a nova), ρ_0 the ambient density, and ξ a dimensionless constant ~ 1 . Once radiative cooling of the shocked material becomes important, the shock is no longer pressure-driven and it will enter the isothermal, momentum conserving phase in which $r_s \propto t^{1/4}$. This point in shock evolution is sometimes referred to as the critical time, t_c , or 'sag time', after which the shock rapidly decelerates.

At time t_c , the dynamical timescale of the shock, t_s , equals the cooling time, t_{cool} . Following the arguments of Bode and Kahn²²

$$t_{\text{cool}} = 0.02 v_s^3 / q \rho_0$$

where v_s is the shock velocity and q ($= 4 \times 10^{32} \text{ cm}^6 \text{ g}^{-1} \text{ s}^{-4}$) is a constant derived from a simple cooling law, for material of solar composition, applicable in the range 10^5 – 3×10^7 K (ref. 23). From the H I measurements, the mean hydrogen density across the region of the nebula is $\geq 25 \text{ cm}^{-3}$. Thus for ejecta from the first nova outburst, encountering an undisturbed cloud, $t_c \sim 170$ yr, at which time $v_s \sim 170 \text{ km s}^{-1}$ and $r_s \sim 2.2 \times 10^{17} \text{ cm}$, that is, an order of magnitude less than the cavity radius implied by the positions of the peaks in the IRAS maps. Note that at this time the shock will have swept up around 10 to 100 times the original mass of the ejecta. Ejecta from the subsequent outbursts will also rapidly decelerate on encountering this swept-up material.

The sputtering timescale is given by:

$$t_{\text{sp}} = \frac{3\rho a}{\mu Y \rho_H v_s}$$

where Y is the sputtering yield, ρ_H the proton mass density, μ the mean molecular weight of the grain material and v_s the shock velocity. For a shock with $v_s = 200 \text{ km s}^{-1}$, $Y \sim 0.01$ (ref. 24), and if $(n_0)_s = 4n_0 = 100 \text{ cm}^{-3}$ in the cloud, then t_{sp} is given for graphite by

$$t_{\text{sp}} = 5,340 \left(\frac{a}{0.1 \mu\text{m}} \right) \left(\frac{(n_0)_s}{100 \text{ cm}^{-3}} \right)^{-1} \left(\frac{v_s}{200 \text{ km s}^{-1}} \right)^{-1} \text{ yr}$$

Similarly, the sweeping-up time is given by

$$t_{\text{sw}} = \rho a / \rho_H v_s$$

giving

$$t_{\text{sw}} = 276 \left(\frac{a}{0.1 \mu\text{m}} \right) \left(\frac{(n_0)_s}{100 \text{ cm}^{-3}} \right)^{-1} \left(\frac{v_s}{200 \text{ km s}^{-1}} \right)^{-1} \text{ yr}$$

Thus both these timescales are in excess of the dynamical timescale at $t = t_c$ so that in addition to shocks being unable to reach the presently observed cavity radius, they would not be effective at sputtering or sweeping up grains to form such a cavity in the dust distribution. In any case, if a cavity has been swept out, we are left with the question of the origin of the medium with which the nova ejecta are now interacting.

As an alternative, one can consider evaporation of grains around the nova at the time of outburst. As $Q_{\text{abs}} \propto \nu^2$ in the far-infrared region (see above) and the grain temperature is low enough that emission is predominantly in this band, it follows that the Planck mean absorption efficiency, $\bar{Q}_{\text{abs}}(T, a) \propto T^2$. From energy-balance considerations, and the results of Draine and Lee²⁵ for $\bar{Q}_{\text{abs}}$, we find

$$r_{\text{evap}} = 46.7 (L \bar{Q}(T_*, a) / T_{\text{evap}}^6 a)^{1/2} \text{ cm}$$

where $\bar{Q}(T_*, a)$ is the Planck mean absorption efficiency at the effective temperature of the nova, L is the nova luminosity, and r_{evap} is the radius of the evaporated cavity with grain evaporation temperature T_{evap} . With $L = 5 \times 10^{38} \text{ erg s}^{-1}$, $\bar{Q}(T_*, a) = 1$, $a = 0.01 \mu\text{m}$, and $T_{\text{evap}} = 1,500 \text{ K}$, $r_{\text{evap}} = 3 \times 10^{14} \text{ cm}$, which is, around a factor of 10^4 smaller than the radius of the cavity. The above conclusions are not altered if silicate grains are considered. If $L \sim 10^{34} \text{ erg s}^{-1}$ at quiescence for GK Per⁸, the temperature of the grains in the toroid, if heated by the quiescent nova radiation field, would only be 12 K. As noted above, the energy source for grain heating is more likely to be the interstellar radiation field. If the grains are indeed similar to those found in the infrared cirrus, then from the correlation between 100 μm surface brightness and A_v (ref. 26) the peak visual optical depth in the cloud is approximately 0.6.

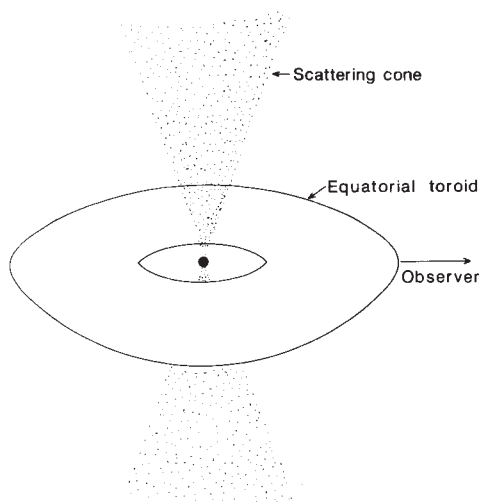


Fig. 2 Proposed model for the morphology of some post main-sequence bipolar nebulae which may also apply to GK Per (after Cohen, ref. 33).

Thus we are inevitably led to conclude that the infrared emission from the cloud is a direct tracer of the mass distribution. In this event, we regard as untenable the hypothesis that the observed infrared emission originates from an interstellar cloud. Its symmetrical displacement about GK Per associates it with a historical ejection of material from the nova. This material could not have been ejected in a previous nova outburst, however, because the mass involved is $\sim 10^4$ – 10^5 times greater than that found in ejected nova envelopes¹.

We propose instead that the morphology of the far-infrared emission is similar to that suggested to account for the bipolar appearance of several planetary nebulae²⁷. Such toroidal structure is assumed to derive from expulsion of the outer layers of a red giant or supergiant star which is a member of a short-period binary system²⁷. Indeed, in the formation of a cataclysmic binary system, planetary nebula ejection is associated with the end of the common envelope phase in which angular momentum is lost from the system²⁸. Support for this view has been given by the detection in recent years of several short-period binaries which are themselves the central stars of planetaries²⁸. In addition, the implied mass of the cloud is not at all unrealistic for the ejected shell of an evolved star, and is in line with the masses derived for the largest planetary nebulae²⁹.

If we are indeed seeing the expanding remnant of the planetary nebula expelled from the central binary of GK Per at the end of its common envelope phase, then it is possible to determine its age. The H I line width shows little evidence for velocities $> 5 \text{ km s}^{-1}$ (E.R.S. *et al.*, in preparation). This low velocity is in line with those seen in the largest planetaries, whose dimensions are $\sim 1 \text{ pc}$ at $z < 150 \text{ pc}$ from the galactic plane (ref. 30, note that GK Per lies at $z < 80 \text{ pc}$). More compact objects have systematically higher expansion velocities, with typical values for young planetary nebulae $\sim 20 \text{ km s}^{-1}$. Deceleration may be a product of the interaction of the expanding envelope with the interstellar medium, or more likely with the product of an earlier mass-loss phase³¹. Thus, with toroidal radius of $\sim 2 \times 10^{18} \text{ cm}$ for a nova distance of 470 pc, an expansion age of between 3×10^4 and $1.3 \times 10^5 \text{ yr}$ results. For a high mass white dwarf, this timescale appears sufficient for a star to decrease in luminosity from a luminous subdwarf to the white-dwarf phase³². (Whether this timescale is sufficient for an isolated white dwarf to reach a luminosity as low as less than a few $L_{\odot}$ (ref. 8) is less clear. However, the true bolometric luminosity of the system is ill-defined, and the luminosity evolution of a post-common-envelope white dwarf is as yet poorly understood. We thank Dr D. Schoenberger for drawing our attention to this point.)

Finally, comparison of the IRAS maps with photographs of

the reflection nebulosity from 1901 and 1902 shows the respective surface brightnesses to be anticoincident (see Fig. 1), highly reminiscent of the proposed morphology of several bipolar nebulae, some of which represent the late stages of stellar evolution. In these objects a dense toroid of material gives rise to infrared emission, and scattering by small dust grains, but with lower overall optical depth, from two cones at right angles to the toroid is inferred (ref. 33, see also Fig. 2). It appears likely that this is the geometry of the nebulosities around GK Persei.

If our model is correct, then the 1901 nova outburst was the first such by this system, hence the interaction of the ejecta with material from an earlier mass loss phase (presumably the high-velocity wind phase which follows the ejection of the main envelope), giving rise to the resultant non-thermal radio emission. That GK Per is a young nova is additionally suggested by its long orbital period and probable eccentric orbit. The density in this inner region is, however, much less than that inferred in the toroid, being $\leq 1 \text{ cm}^{-3}$ (ref. 13). The placement of the reflection nebulosity suggests that this flow continues to much greater distances from the nova.

In conclusion, the apparent discovery of an ancient planetary nebula around GK Per has many important ramifications for our understanding of the evolutionary status of nova binaries, and also the later phases of planetary nebula development. In addition, we may be close to unravelling the 86-year-old mystery of why GK Persei appears so different in so many respects from other classical novae.

We thank Dr M. Barlow, Dr H. W. Duerbeck, Mr D. H. Hughes and Professor W. Seitter for stimulating discussions; Drs P. Richards and B. Stewart for help with the analysis of IRAS data, and Pat Shand and Diana Larson of the Lick Photo Labs for provision of archival materials. M.F.B. and J.S.A. are supported by the UK Science and Engineering Research Council; J.A.R. is supported by the National Advisory Body to the Polytechnics; E.R.S. acknowledges the support of an operating grant from the Natural Sciences and Engineering Research Council of Canada (NSERC); D.A.F. was supported by an NSERC scholarship.

Received 7 April; accepted 15 July 1987.

1. Bath, G. T. *Q. Jl R. astr. Soc.* **19**, 442–455 (1978).
2. Sabbadin, F. & Bianchini, A. *Astr. Astrophys. Suppl. Ser.* **54**, 393–403 (1983).
3. Schweizer, E. *IAU Telegram* no. 4274 (1986).
4. Bianchini, A., Hamzaoglu, E. & Sabbadin, F. *Astr. Astrophys.* **99**, 392–393 (1981).
5. Crampton, D., Cowley, A. P. & Fisher, W. A. *Astrophys. J.* **300**, 788–793 (1986).
6. Watson, M. G., King, A. R. & Osborne, J. *Mon. Not. R. astr. Soc.* **212**, 917–930 (1985).
7. Crampton, D., Cowley, A. P. & Hutchings, J. B. in *Cataclysmic Variables and Related Objects* (eds Livio, M. & Shaviv, G.) 25–28 (Reidel, Dordrecht, 1983).
8. Bianchini, A. & Sabbadin, F. *Astr. Astrophys.* **125**, 112–116 (1983).
9. Perrine, C. D. *Astrophys. J.* **15**, 136–141 (1902).
10. Couderc, P. *Ann. Astrophys.* **2**, 271–302 (1939).
11. Swope, H. H. *Harv. Obs. Bull.* no. 913, 11 (1940).
12. Van den Bergh, S. *Publ. Astr. Soc. Pacif.* **89**, 637 (1977).
13. Reynolds, S. & Chevalier, R. *Astrophys. J. Lett.* **281**, L33–L35 (1984).
14. Bode, M. F., Seaquist, E. R. & Evans, A. *Mon. Not. R. astr. Soc.* **227**, 217–227 (1987).
15. Beichman, C. A., Neugebauer, G., Habing, H. J., Clegg, P. E. & Chester, T. J. (eds) *IRAS Explanatory Suppl.* (Jet Propulsion Lab. Publ. JPL, D-1855, 1985).
16. Low, F. J. *et al. Astrophys. J.* **278**, L19–L22 (1984).
17. Draine, B. T. *Astrophys. J. Suppl. Ser.* **57**, 587–594 (1985).
18. Weiland, J. L. *et al. Astrophys. J.* **306**, L101–L104 (1986).
19. McLaughlin, D. B. in *Stars and Stellar Systems* Vol. 6 (ed. Greenstein, J. L.) 585–652 (University of Chicago Press, 1960).
20. Savage, B. D. & Mathis, J. S. *A. Rev. Astr. Astrophys.* **17**, 73–111 (1979).
21. Sedov, L. I. in *Similarity and Dimensional Methods in Mechanics* (Academic, New York, 1959).
22. Bode, M. F. & Kahn, F. D. *Mon. Not. R. astr. Soc.* **217**, 205–215 (1985).
23. Kahn, F. D. *Astr. Astrophys.* **50**, 145–148 (1976).
24. Draine, B. T. & Salpeter, E. E. *Astrophys. J.* **231**, 77–94 (1979).
25. Draine, B. T. & Lee, H. M. *Astrophys. J.* **285**, 89–108 (1984).
26. De Vries, C. P. & Le Poole, R. S. *Astr. Astrophys.* **145**, L7–L9 (1985).
27. Balick, B. in *Proc. Workshop on the Late Stages of Stellar Evolution* (ed. Kwok, S.) (D. Reidel, Dordrecht, in the press).
28. Eggleton, P. in *Cataclysmic Variables and Related Objects* (eds Livio, M. & Shaviv, G.) 239–255 (Reidel, Dordrecht, 1983).
29. Pottasch, S. R. *Astr. Astrophys.* **89**, 336–341 (1980).
30. Geiseking, F., Hippelein, H. & Weinberger, R. *Astr. Astrophys.* **156**, 101–105 (1986).
31. Sabbadin, F., Gratton, R. G., Bianchini, A. & Ortolani, S. *Astr. Astrophys.* **136**, 181–192 (1984).
32. Iben, I. & Renzini, A. *A. Rev. Astr. Astrophys.* **21**, 271–342 (1983).
33. Cohen, M. in *Planetary Nebulae* (ed. Flower, D. R.) 45–55 (Reidel, Dordrecht, 1983).

Water frost on Charon

Marc W. Buie*, Dale P. Cruikshank*,
Larry A. Lebofsky† & Edward F. Tedesco‡

* Institute for Astronomy 2680 Woodlawn Drive, Honolulu, Hawaii 96822, USA

† Lunar and Planetary Laboratory, University of Arizona, Tucson, Arizona 85721, USA

‡ Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California 91109, USA

The current series of mutual eclipses between Pluto and its satellite, Charon, provides a very powerful means of probing the most distant known planet in our Solar System. Observations from 1985 and 1986 have already dramatically improved our knowledge of the sizes, albedos, and the orbital parameters of the system¹⁻³. One experiment that we had been waiting to perform is to observe Pluto during a total eclipse of its satellite. This geometry provides a direct means to study the planet without contamination from the satellite. Once the spectrum of the planet is known, it is then possible to subtract it from the spectrum of the planet plus satellite and thus discern the properties of the satellite. Here we present new spectra of the Pluto-Charon system taken just before and during a total eclipse of the satellite. From these data we have extracted the spectrum of the satellite, Charon, which reveals the spectral signature of water ice. There is no evidence for any methane or ammonia frost on the surface of Charon. This observation places important constraints on the composition and origin of this planetary system.

On 3 March 1987 UT Marcialis *et al.* obtained observations⁴ of a total eclipse of the satellite at four discrete wavelengths, 1.5, 1.75, 2.0 and 2.35 μm . Their observations provided a tentative identification of water ice on Charon through the excellent placement of the chosen wavelengths. Our efforts at Mauna Kea on the same date were hampered by high winds and we could not provide simultaneous observations.

The observations reported here were made by one of us (M.W.B.) at the IRTF (infrared telescope facility) on Mauna Kea, Hawaii, on the night of 23 April 1987 UT. The detector was an InSb diode used in a spectrometer in which a circular variable interference filter (CVF) provided the wavelength separation. The sensitivity of this system was such that we could obtain a CVF spectrum of Pluto with a signal-to-noise ratio of ~ 100 in one hour.

We used an 8-arc s aperture with a 5% resolution, high-throughput CVF covering the wavelength range 1.5–2.5 μm . A single spectrum consists of 13 different points equally spaced across this wavelength range. With the resolution of this CVF a 13-point spectrum is slightly undersampled. The chopping distance was 20 arc s to the north, with a chopping frequency of 6 Hz. The comparison star for these observations was SAO120599. This G0 star should match the near-infrared spectrum of the Sun very closely so that a ratio spectrum of Pluto to the star will be a close approximation of its spectral reflectance. The same star was used by Marcialis *et al.*⁴.

The predicted event timing was 11:13 UT for first contact, 13:07 UT for second contact, 13:59 UT for third contact, and 15:44 UT for fourth contact⁵. It is between second and third contact that the satellite is completely hidden from view. Before the event, D. J. Tholen provided an updated prediction that second contact would be ten minutes early and third contact five minutes late.

The out-of-eclipse spectrum was started at 9:20:09 UT and completed at 10:20:52 UT. The spectrum during totality was taken between 13:00:35 UT and 14:00:04 UT, thus matching the predicted geometry very closely. In all cases, a spectrum of the comparison star was taken before and after each spectrum of Pluto. Throughout the observations the tracking of the telescope

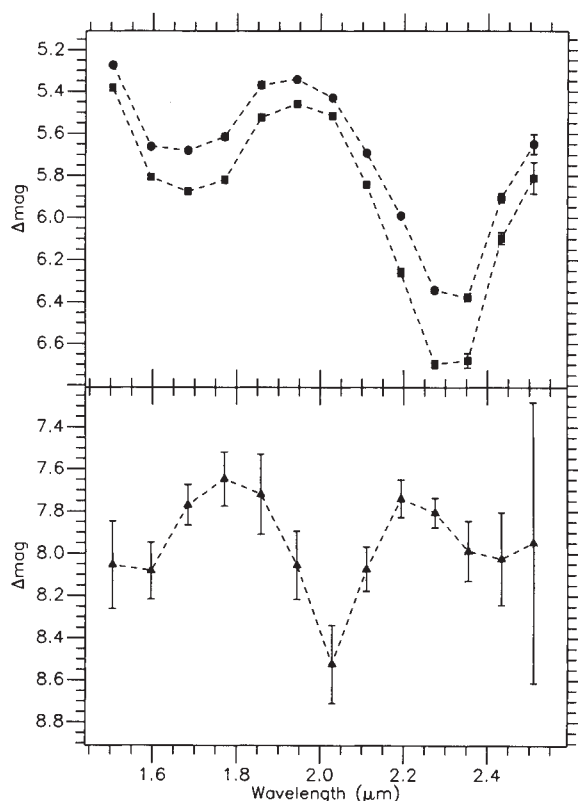


Fig. 1 *a*, Magnitude plot of spectra of the Pluto-Charon system before and during the total eclipse of the satellite. *b*, Spectrum of Charon derived from subtracting the two spectra in *a* after correcting for lightcurve effects. The stellar magnitudes are relative to SAO120599. ●, Pluto and Charon; ■, Pluto, ▼, Charon.

on Pluto was checked about every five or ten minutes to ensure that there were no systematic drifts of Pluto out of the aperture.

The fluxes were recorded in units of stellar magnitude, and the instrumental magnitudes for each point in the spectrum were converted to colours by subtracting each magnitude in a given spectrum from a reference wavelength, 2.03 μm . The colour data were then reduced as standard photometry using the comparison star to provide the extinction correction at each wavelength. By reducing the data as colours, errors in the extinction correction were minimized. The colour spectra were then converted back to magnitude spectra for analysis by adding the magnitude at the reference wavelength. The errors derived are in units of a standard deviation.

Since the comparison star is similar in colour to the Sun, the reduced spectra should closely represent the spectral reflectance of the system. Figure 1*a* shows the out-of-eclipse spectrum of Pluto and Charon together with the eclipse of Pluto alone. Strong absorptions due to methane, centred at 1.7 and 2.3 μm , dominate both spectra. The error bars shown for each point reflect the known random errors from all stages of the data reduction. The signal-to-noise ratio is ~ 100 , representing the highest signal-to-noise spectra of this system ever published for this wavelength range. These spectra are plotted on a magnitude scale so that any changes in brightness will not affect the shape of the spectral features. Any change in the shape of the spectrum on this plot is due entirely to the loss of the light from Charon. Note that the methane bands are clearly deeper when the light from Charon is removed.

Before the spectrum of Charon can be extracted, the effect of the underlying lightcurve of the system must be removed. At this rotational phase the slope of the background lightcurve of the system is 0.0055 mag per hour (D. J. Tholen and E.F.T., in preparation; M.W.B. and D. J. Tholen, in preparation). In the 3.67-h interval between taking the spectra, the system faded by

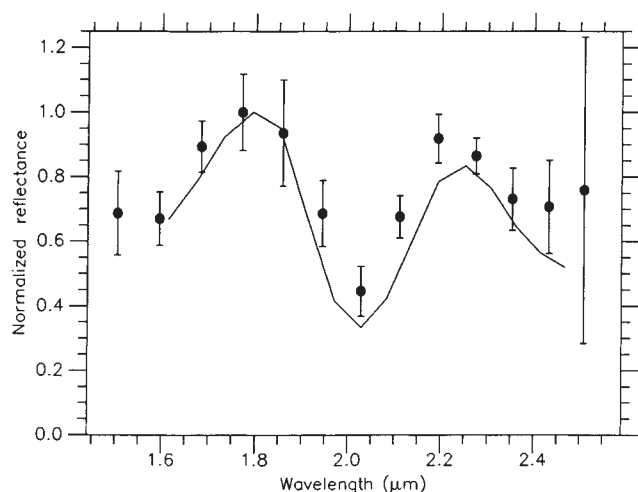


Fig. 2 Infrared spectrum of Charon compared to a laboratory water frost. The solid curve is a laboratory spectrum of a medium-fine-grained H_2O frost. The plot shows the data from Fig. 1a as relative reflectance, normalized to 1 at $1.8\ \mu\text{m}$.

0.020 mag. Here we have assumed that all of the lightcurve slope is due to Pluto. Therefore the out-of-eclipse spectrum was reduced by 0.020 mag before subtracting the Pluto spectrum. As Pluto contributes most of the light ($\sim 80\%$), we would expect most of the slope to be due to Pluto; though the precise contributions of Pluto and Charon to the lightcurve are not known. Furthermore, the slope of the lightcurve is well determined only at visual wavelengths. By using the visual lightcurve slope, we are extrapolating the lightcurve behaviour from 0.4 to $2.0\ \mu\text{m}$, on the assumption that the slope is independent of colour. As it has already been proved that the methane spectrum of Pluto varies with rotational phase^{6,7}, we know that this assumption is flawed.

Figure 1b shows the spectrum of Charon obtained by subtracting the two spectra shown in Fig. 1a after the lightcurve correction is applied. The characteristic $2\text{-}\mu\text{m}$ water frost absorption feature seen in the spectrum of Charon is unmistakable. Shown for comparison in Fig. 2 is a laboratory spectrum of a medium-fine-grained H_2O frost⁸. The spectrum of Charon is very different from that of Pluto alone, as can be seen in Fig. 1. In fact, the lack of any residual spectral signature due to methane (the spectrum of Pluto) gives us confidence that the lightcurve correction is accurate.

The spectrum of Charon is quite similar to those of other satellites in the outer solar system. Europa, Ganymede, Callisto, most of the saturnian satellites, and the uranian satellites all have spectra dominated by water frost absorptions⁹. The spectrum of Miranda is perhaps the closest match¹⁰.

The most interesting result is that the spectra of Pluto and Charon are so different. If both bodies have the same origin, then the methane Charon originally had on its surface must have sublimated and escaped. In addition, the lack of a water frost signature on Pluto implies that any water frost there has been covered with methane. Both the lower albedo of Charon relative to Pluto and the lower surface gravity could have combined over the age of the solar system to remove all traces of methane from at least the near surface layers on Charon. If this is the case, we can now begin to set some limits on the composition and origins of the two bodies. The difference in the surface composition makes the measurement of the relative densities within the system much more important. Fortunately, this is a problem within the range of the Hubble Space Telescope.

We thank D. Toomey for his work on Primo the exceptional detector with which these observations were made. Thanks also go to D. Morrison, A. Tokunaga, and R. Koehler, for their continued support of this difficult observational project and to D. Tholen for updated eclipse predictions. The authors were

Visiting Astronomers at the Infrared Telescope Facility, which is operated by the University of Hawaii, under contract with NASA. We thank NASA for support of this research.

Received 9 June; accepted 29 July 1987.

1. Tholen, D. J., Buie, M. W., Binzel, R. P. & Fruch, M. *Science* **237**, 512–514 (1987).
2. Reinsch, K. & Pakull, M. W. *Astr. Astrophys.* **177**, L43 (1987).
3. Dunbar, R. S. & Tedesco, E. F. *Astr. J.* **92**, 1201–1209 (1986).
4. Marcialis, R., Reike, G. & Lebofsky, L. *Science* (in the press).
5. Tholen, D. J., Buie, M. W. & Swift, C. E. *Astr. J.* **93**, 244–247 (1987).
6. Buie, M. W. & Fink, U. *Icarus* **70**, 483–498 (1987).
7. Cruikshank, D. P. & Brown, R. H. in *Ices in the Solar System* (eds Klinger, J., Benest, D., Dollfus, A. & Smoluchowski, R.) 817–827 (Reidel, Dordrecht, 1985).
8. Clark, R. N. *J. geophys. Res.* **86**, 3087–3096 (1981).
9. Clark, R. N., Fanale, F. P. & Gaffey, M. J. in *Ices in the Solar System* (eds Klinger, J., Benest, D., Dollfus, A. & Smoluchowski, R.) 437–491 (Reidel, Dordrecht, 1985).
10. Brown, R. H. & Clark, R. N. *Icarus* **58**, 288–292 (1984).

Determination of the magnetic penetration depth of the high- T_c superconductor $\text{YBa}_2\text{Cu}_3\text{O}_{7-x}$ by polarized neutron reflection

R. Felici[‡], J. Penfold^{*}, R. C. Ward^{*}, E. Ols[†] & C. Maticotta[†]

^{*} Rutherford Appleton Laboratory, Chilton, Oxfordshire OX11 0QX, UK

[†] ITM-CNR, Via Induno 10, 20091 Cinisello Balsamo, Italy

The recent discovery of high-temperature superconductivity in $\text{La}_{1.85}\text{Ba}_{0.15}\text{CuO}_4$ ¹ and $\text{YBa}_2\text{Cu}_3\text{O}_{7-x}$ ², with critical temperatures T_c of $\sim 40\ \text{K}$ and $> 90\ \text{K}$ respectively, has resulted in great activity in the investigation of these materials. Associated with the Meissner effect is the well-known surface flux penetration; a penetration depth of $\sim 1,500\ \text{\AA}$ has been derived³ for $\text{YBa}_2\text{Cu}_3\text{O}_{7-x}$, and values of $\sim 2,000\ \text{\AA}$ were measured by muon spin rotation^{4,5} for $\text{La}_{1.85}\text{Ba}_{0.15}\text{CuO}_4$. Felcher and co-workers^{6,7}, in their work on superconducting films of niobium, lead and lead-bismuth, have used the reflection of spin-polarized slow neutrons to obtain a direct and absolute measurement of the penetration depth. We report here the first such measurement of the penetration depth in a sample of $\text{YBa}_2\text{Cu}_3\text{O}_{7-x}$. At a temperature of $4.8\ \text{K}$ and in an applied magnetic field of $350\ \text{oersteds}$ we obtain a value, which represents an upper limit, of $225 \pm 75\ \text{\AA}$, which is small compared with penetration depths in conventional superconductors⁸, and with the recently quoted values for $\text{YBa}_2\text{Cu}_3\text{O}_{7-x}$ ³ and $\text{La}_{1.85}\text{Ba}_{0.15}\text{CuO}_4$ ^{4,5}.

The reflection of X rays⁹ and neutrons^{10,11} has been used extensively to study the composition of materials at surfaces and interfaces. The reflection of slow neutrons can be described by the standard laws of optics¹², the reflectivity being described in terms of the refractive index $n(z)$ in the material as a function of the distance z from the surface. For most materials the refractive index is less than unity: materials are optically less dense than vacuum, and total external reflection takes place at an air-material interface. For magnetic materials the refractive index has a magnetic contribution in addition to the nuclear one. For materials magnetized parallel to the surface, neutrons polarized parallel (+) or antiparallel (−) to the direction of the applied magnetic field H have a spin-dependent refractive index

$$n^\pm(z) = 1 - \frac{\lambda^2}{2\pi} \left[\frac{b}{V} \pm C(B(z) - H) \right] \quad (1)$$

where λ is the neutron wavelength, b is the average nuclear scattering amplitude in an atomic volume V , C is a constant equal to $2.3 \times 10^{-10}\ \text{\AA}^{-2}\ \text{Oe}^{-1}$ and $B(z)$ is the magnetic induction

[‡] Permanent address: ISMA-CNR, Via E. Fermi, 00044 Frascati, Italy.

in the material. A measurement of the reflectivity in the region of total reflection as a function of wavevector transfer K (where $K = 4\pi \sin \theta / \lambda$ and θ is the glancing reflection angle) perpendicular to the surface allows a determination of $n(z)$ and, in the case of a magnetic material, measurement of the spin-dependent reflectivities R_+ and R_- allows a unique determination of the surface depth profile of the magnetization $B(z)$. Total reflection of spin-polarized slow neutrons has already been used to study magnetization in thin films¹³ and to determine the penetration depth in conventional superconductors^{6,7}.

The sample was prepared from powders of Y_2O_3 , BaCO_3 and CuO to the correct composition, and the mixture was reacted in flowing oxygen in a silica crucible at 950 °C for 16 hours. The powdered sample was compacted at a pressure of 10 kbar to a disc 16 mm in diameter and 3 mm thick, and sintered at 950 °C in flowing oxygen. One surface was dry-polished using a diamond disk to produce a smooth finish for the neutron total reflection measurements. Electron micrographs of the polished surface at low magnification ($\times 1,000$) show surface regions of high density (with characteristic dimensions of $\sim 10 \mu\text{m}$) separated by smaller voids which occupy $\sim 30\%$ of the surface. At higher magnification ($\times \sim 10,000$) we observe the granular nature of the sample and obtain a characteristic dimension of the individual grains of $\sim 2 \mu\text{m}$. Resistivity measurements were made on an identical sample, yielding a critical temperature of 91.2 K with a width of 2.2 K. Measurements of the total reflection from the polished surface were made using the polarized and unpolarized modes of the reflectometer CRISP¹⁴ at the ISIS facility of the Rutherford Appleton Laboratory. The measurements were made as a function of neutron wavelength (measured by time of flight) in the range 2–6.5 Å, using a fixed glancing angle of incidence $\theta = 0.3^\circ$, and a beam divergence of $\pm 0.006^\circ$.

Figure 1 shows the spin-independent reflectivity measured at room temperature and with no applied magnetic field. The solid line is a fit to the data for $b/V = 3.4 \times 10^{-6} \text{ Å}^{-2}$. The reflectivity calculation (solid line in Fig. 1) is averaged over a gaussian angular spread of 25% in $\Delta\theta/\theta$. This includes the divergence of the incident beam ($\sim 3.7\%$ in $\Delta\theta/\theta$), but is dominated by surface undulations. Locally the surface is not smooth, and its roughness can be considered to modify the reflected intensity as a Debye–Waller factor¹⁵:

$$I(\lambda) = I_0(\lambda) \exp(-16\pi^2 \sin \theta \sin \theta_1 \langle z_D^2 \rangle / \lambda^2) \quad (2)$$

where θ_1 is the refracted angle and z_D is the root-mean-square roughness. The roughness parameter has been included in the calculation, and gave a value of $z_D = 90 \text{ Å}$, similar to that obtained for the sputtered niobium film in ref. 6. The characteristic dimension of the high-density regions observed in the electron micrographs is less than the length scale over which phase coherence is maintained in the total reflection measurement. The reflectivity from the surface in such circumstances is identical to that from a homogeneous surface of lower scattering density; this therefore explains the difference between the fitted value of b/V of $3.4 \times 10^{-6} \text{ Å}^{-2}$, and that of $4.2 \times 10^{-6} \text{ Å}^{-2}$ expected from crystallographic data¹⁶.

The spin-dependent reflectivities R_+ and R_- were measured under experimental conditions identical to those for the data in Fig. 1, at a temperature of 4.8 K for applied magnetic fields of 350, 715, 1,100 and 1,300 Oe, and at 130 K (above T_c) for an applied field of 705 Oe. The sensitivity of the technique to the magnetic component of the refractive index is most clearly demonstrated in the spin reflectivity ratio R_+/R_- , and the data in Fig. 2 are presented in this form for the different applied magnetic fields and temperatures. Above T_c the spin reflectivity ratio is unity (Fig. 2d), indicating that there is no diamagnetism within the material. The lines in Fig. 2 are calculated curves, for which we have assumed a complete Meissner effect (meaning that the bulk material is fully diamagnetic) and a finite magnetic induction in the surface of the form $B(z) = H \exp(-z/\Lambda)$, where z is the distance from the surface, and Λ is the magnetic-field

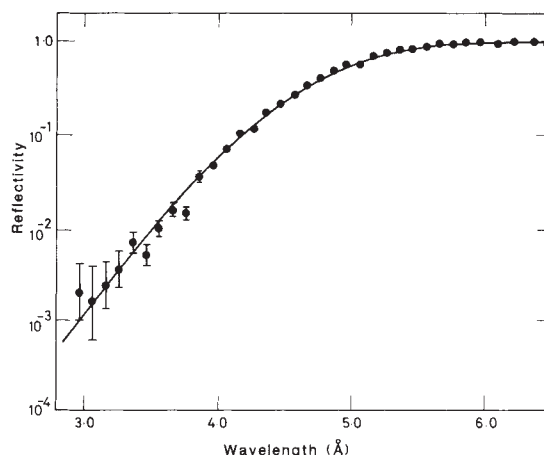


Fig. 1 Reflectivity of $\text{YBa}_2\text{Cu}_3\text{O}_{7-x}$ at an angle of incidence $\theta = 0.3^\circ$. The solid line is a fit to the data with a gaussian spread of angles $\Delta\theta = 25\%$, and a surface roughness parameter $\langle z_D^2 \rangle^{1/2} = 90 \text{ Å}$.

penetration depth. The values of nuclear scattering density, surface roughness and angular spread of the reflected beam are taken from the fit to the data in Fig. 1. The solid lines are obtained for penetration depths of 150 and 300 Å for the lower and upper curves respectively. The dashed line is for a penetration depth of 2,000 Å, corresponding to the value obtained from the muon spin rotation (μSR) measurement of ref. 4 for $\text{La}_{1.85}\text{Ba}_{0.15}\text{CuO}_4$. From our data we estimate that the magnetic penetration depth is $225 \pm 75 \text{ Å}$. This is very different from other reported measurements. Muon spin rotation measurements on $\text{La}_{1.85}\text{Ba}_{0.15}\text{CuO}_4$ ^{4,5,17}, magnetization studies on $\text{La}_{1.85}\text{Ba}_{0.15}\text{CuO}_4$ ¹⁸ and $\text{YBa}_2\text{Cu}_3\text{O}_{7-x}$ ³ and infrared reflectance measurements¹⁹ all give values of $\sim 2,000 \text{ Å}$; in contrast, a neutron small-angle scattering measurement²⁰ has yielded a value ($\sim 21 \text{ Å}$) even smaller than that reported here.

The value of Λ quoted by us refers to applied fields $H \leq H_{cl}$, but we have measured the field dependence of the spin reflectivity ratio (see Fig. 2b, c) and we observe no change for fields $\geq 715 \text{ Oe}$, consistent with being at an applied field $H \geq H_{cl}$. The calculated line in Fig. 2b is for an applied field of 715 Oe and a penetration depth of 225 Å. As the field is $\geq H_{cl}$ there is a finite flux inside the sample and our initial assumptions no longer apply. But the degree of agreement between the data and calculated curve for $H = 715 \text{ Oe}$ suggests that H is quite close to H_{cl} , and we estimate H_{cl} to be $600 \pm 100 \text{ Oe}$. This is further supported by the even larger disagreement between the data and calculation at the higher applied field of 1,100 Oe (Fig. 2c).

The measurement of Λ is an average over all crystallographic directions because of the polycrystalline nature of the sample; there is insufficient information with which to deduce any anisotropy in the penetration depth with respect to lattice direction.

If we consider the nature of the sample and its surface, then the value of Λ quoted should be regarded as an upper limit for the following reasons:

(1) We have assumed a total Meissner effect, and if the diamagnetism is not perfect then a shorter penetration depth would be required to produce the same effect. Assuming a Meissner effect of 85%, a penetration depth of $\sim 100 \text{ Å}$ is required to fit the data. Although the magnitude of the effect can be predicted for imperfect diamagnetism with a reduced penetration depth, it appears increasingly difficult, for a Meissner effect of $< 85\%$, to preserve good agreement with the shape of the spin reflectivity curve.

(2) With regard to the granular nature of the sample, we have observed from electron microscopy that the grain size is $\sim 2 \mu\text{m}$, which is much greater than our measured penetration depth. Furthermore, our measurement is an integration over many

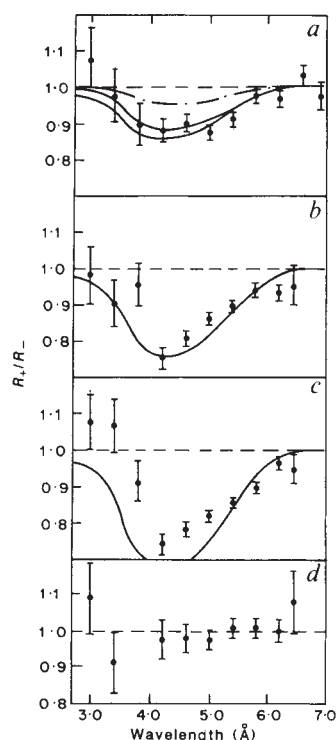


Fig. 2 Spin reflectivity ratio R_+/R_- , for $\theta = 0.3^\circ$, $T = 4.8$ K (a–c) and 130 K (d). The externally applied field H is: a, 350; b, 715; c, 1,100; d, 705 Oe. Solid lines in a are calculations for $\Lambda = 300$ Å (upper curve) and 150 Å (lower curve); dashed line is for $\Lambda = 2,000$ Å. Solid lines in b and c are calculations for $\Lambda = 225$ Å.

grains, and we have discussed above the effect of this on the surface reflectivity. An additional consequence of the granular nature is that magnetic flux may exist within the material. This is tantamount to the first point about incomplete diamagnetism, and because of the nature of our averaging any complex form of the flux pattern will not be significant in calculating its effect.

(3) If the surface structure is modified by the polishing then superconductivity could be lost at the surface. Calculations incorporating a dead layer at the surface again show that a shorter penetration depth would result (even a ferromagnetic surface layer has qualitatively the same effect).

(4) The surface roughness is of the same order of magnitude as the magnetic penetration depth. However we have only used this parameter as an optical roughness modifying the expression for the spin-dependent reflectivities in the form of a Debye–Waller factor¹⁵. If the roughness implies a modification to the surface structure then again our value of penetration depth is an upper limit, as in point (3).

We are left with a substantial discrepancy between our measurement of the penetration depth and other observations^{4,5,17,18,20}. We have discussed factors which would alter our interpretation towards an even shorter penetration depth; this would be more consistent with recent small-angle neutron scattering measurements²⁰ than with the other observations predicting longer penetration depths^{4,5,17,18}. Previous measurements of penetration depth by spin-polarized slow neutrons⁶ for conventional superconductors are in agreement with measurements by other techniques. The spin-polarized slow-neutron reflection method is direct and does not suffer from the complications and implicit assumptions inherent in the other methods. The μ SR measurements, for example, have been made in fields above and below H_{c1} and their interpretation is based on formulae derived from the regular flux lattice found in conventional type 2 superconductors above H_{c1} , and assumes that the London theory is applicable. Limitations of this important assumption have been observed for niobium²¹. The macro-

scopic measurements are even more indirect and make assumptions based on particular theoretical aspects of conventional superconductivity.

Received 16 June; accepted 12 September 1987.

1. Bednorz, J. G. & Müller, K. A. *Z. Phys.* **B64**, 189–193 (1986).
2. Wu, M. K. *et al.* *Phys. Rev. Lett.* **58**, 908–910 (1987).
3. Cava, R. J. *et al.* *Phys. Rev. Lett.* **58**, 1676–1679 (1987).
4. Kossler, W. J. *et al.* *Phys. Rev.* **B35**, 7133–7134 (1987).
5. Aeppli, G. *et al.* *Phys. Rev.* **B35**, 7129–7132 (1987).
6. Felcher, G. P., Kampwirth, R. T., Gray, K. E. & Felici, R. *Phys. Rev. Lett.* **52**, 1539–1542 (1984).
7. Felcher, G. P., Felici, R., Kampwirth, R. T. & Gray, K. E. *J. appl. Phys.* **57**, 3789–3794 (1985).
8. Tinkham, M. *Introduction to Superconductivity* (McGraw-Hill, New York, 1975).
9. Parrat, L. G. *Phys. Rev.* **95**, 359–369 (1954).
10. Hayter, J. B. *et al.* *JCS Faraday Trans. I* **77**, 1437–1448 (1981).
11. Highfield, R. R. *et al.* *Thin Solid Films* **99**, 165–179 (1983).
12. Born, M. & Wolf, E. *Principles of Optics* (Pergamon, Oxford, 1970).
13. Bland, J. A. C., Pesca, D. & Willis, R. F. *Phys. Rev. Lett.* **58**, 1244–1247 (1987).
14. Penfold, J., Ward, R. C. & Williams, W. G. *J. Phys. E* (in the press).
15. Névoit, L. & Croce, P. *Rev. Phys. Appl.* **15**, 761–779 (1980).
16. David, W. I. F. *et al.* *Nature* **327**, 310–312 (1987).
17. Wappling, R. *et al.* *Phys. Lett.* **A122**, 209–214 (1987).
18. Finnemore, D. K. *Phys. Rev.* **B35**, 5319–5321 (1987).
19. Tajima, S., Uchida, S., Kaube, S., Kitizawa, K. & Fueki, K. *Jap. J. appl. Phys.* (in the press).
20. Lynn, J. *et al.* *Phys. Rev. B* (in the press).
21. Schelten, J., Ullmaier, H., Lippmann, G. & Schmatz, W. *Proc. 13th int. Conf. Low Temp. Phys.* Vol. 3, 54–63 (North Holland, Amsterdam, 1974).

An ordering transition in the turbulent flow of helium II

L. C. Krysac & G. M. Graham

Department of Physics, University of Toronto, Toronto, Ontario, M5S 1A7, Canada

We report the observation of some remarkable features in the flow of a vertical, planar jet of superfluid helium forced through a macroscopic orifice while carrying heat. We observed a transition from a state in which the profile of the free surface of the jet carried a Kelvin-like wave to a state with an almost undisturbed interface. This transition was observed as a heat flux in the jet stream was increased. This new state presented a bifurcation of the stream at the orifice such that a second branch coursed through a narrow stream into a hydraulic jump to the side of the jet and orifice. The hydraulic jump exhibited sharply periodic surface waves travelling away from the orifice. The stream velocities and heat currents for which this new state was observed greatly exceeded the critical velocity¹ and the critical heat current² for the onset of quantum turbulence. This sudden transition into a less chaotic steady-flow configuration above velocities large enough to induce shearing instabilities in a stream has no precedent in the hydrodynamics of ordinary fluids and is a phenomenon which has not been previously observed in superfluid helium.

In this experiment the superfluid flow was constrained to a plane-parallel geometry by two sheets of clear Lexan plastic. The orifice and flow channel of the jet were formed by carving 1-mm-thick pieces of the same plastic and bonding them between the two thicker sheets. Either gravity or the thermomechanical effect³ could be used to drive the flow. When the thermomechanical effect was used, ~ 0.5 Ω of chromel heater wire was inserted into a chamber above a superleak of tightly packed 0.05 μ m Alpha alumina polishing powder. The bottom of the chamber was submerged in a bath of He II and the top of the chamber formed a sharply angled orifice through which the jet emerged into the helium vapour above the reservoir. When gravity was employed, the superfluid contained in a reservoir was allowed to drain through a channel which directed the stream in a vertical direction through a similar orifice. A 0.5- Ω heater wire was inserted in the channel so that the effect of the heat current supplied by the heater could be observed. In both types of fountains the orifice widths ranged from 1 mm to 3 mm.

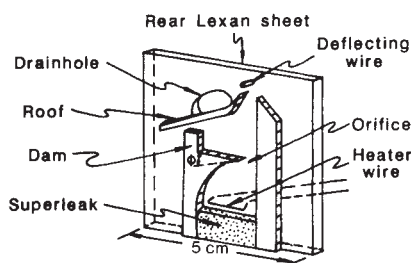


Fig. 1 A scaled drawing of the plane-parallel-geometry flow chamber, with the front Lexan sheet omitted for clarity. The attack angle ϕ of the orifice is indicated.

The ordering transition observed occurred only when the following elements were present in the experiment: a stream of superfluid with relatively high velocity ($>50 \text{ cm s}^{-1}$) flowing past a sharply angled orifice with a relatively large heat current in the stream ($>1.5 \text{ mW}$). Experiments with the gravity flow showed that the transition does not occur at temperatures above the Lambda point. The transition has not been observed in experiments with a three-dimensional circular orifice.

The geometry illustrated in Figs 1 and 2a-d is for a thermomechanically driven jet forced through a one-sided orifice. A roof and drain are inserted above and to one side of the jet to allow the jet to drain away without interfering with the flow of superfluid near the orifice. A wire which spanned the Lexan sheets is inserted near the lip of the roof such that it diverted the jet to one side. All the phenomena described here occurred whether or not these inserts were present. The 5-mm-high dam placed at $\sim 1.5 \text{ cm}$ to the left of the orifice enhanced the development of a hydraulic jump, which shall be described later. As the velocity of the jet at the orifice, given by $v^2 = 2gH$ (where g is the acceleration of gravity, and H is the jet height above the orifice), and the heat current were increased, the flow pattern of the liquid helium developed through several stages. As the superfluid inside the fountain chamber was heated, a pressure head arose due to the thermomechanical effect. This head eventually emerged from the orifice, spilled over the dam to the left of the orifice, and developed into a jet and a small pool of liquid abutted against the dam. (Fig. 2a, b). At this point, the velocity of the fluid at the orifice was $\sim 60 \text{ cm s}^{-1}$. For stream velocities between this and $\sim 70 \text{ cm s}^{-1}$, the free vertical surface of the jet was distorted by aperiodic waves, accompanied by liquid whiskers projecting from the jet at an angle of $\sim 160^\circ$ to the streaming direction. The form of the aperiodic waves and streaks was temperature dependent, acquiring the familiar form of Kelvin-Helmholtz waves⁴ at higher temperatures near 1.6 K (Fig. 2c, d).

At a velocity near 60 cm s^{-1} a sudden transition took place. A solitary wave was observed on the surface of the pool, moving away from the barrier towards the jet, (Fig. 2e, f), such that, when it arrived at the jet, the jet and pool divided into two bodies of fluid separated by a thin region of vapour. The sustained pool became a hydraulic jump which did not collapse into the jet despite its apparent opportunity to do so. At the exact moment of this spatial bifurcation, the surface of the jet became dramatically quiescent, exhibiting only small, barely visible wave disturbances on its surface. The exact velocity of the jet stream when this reversible transition took place was sensitive to the angle of the orifice and the heat flux, as discussed below.

Increasing the velocity of the jet from 70 to 100 cm s^{-1} resulted in periodic waves on the surface of the hydraulic jump, travelling away from the orifice, with increasing frequency and amplitude for increasing jet velocity. (Fig. 2g, h). The observed frequencies, given by the frequency at which the waves appeared stationary when observed under stroboscopic lighting, ranged from 30 to 80 Hz, for different orifices and jet velocities. The amplitude of

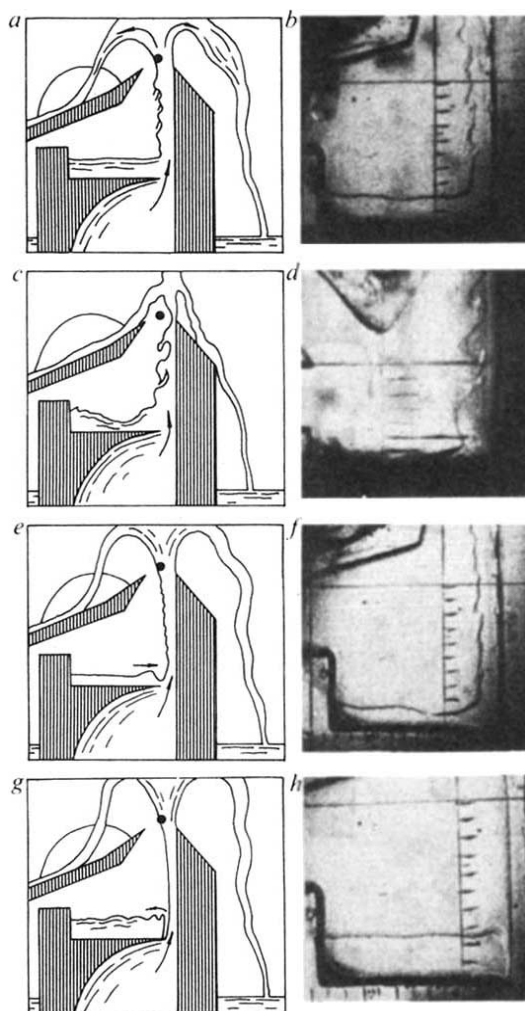


Fig. 2 Photographs and schematic drawings showing the development of the superfluid flow as the stream velocity past the orifice was increased. The photographs are not on the same scale as the drawings. The horizontal dashes in the photographs are 1 mm apart. The approximate velocity ranges were: a, b, from 60 to 70 cm s^{-1} ; c, d, 65 cm s^{-1} at a higher temperature 1.6 K; e, f, near 70 cm s^{-1} ; g, h, from 70 to 100 cm s^{-1} .

the surface wave was $\sim 1 \text{ mm}$. The depth of the hydraulic jump extended to the tip of the orifice as far as we could discern.

At a jet velocity of $\sim 100 \text{ cm s}^{-1}$ and a hydraulic jump height of $\sim 3 \text{ mm}$, the periodicity of the surface waves broke down and their motion became increasingly agitated, developing into elongated peaks which rounded the crest of the hydraulic jump like spokes on a wheel, eventually appearing like long whips of fluid snapping out into the vapour (Fig. 3). At a jet velocity of $\sim 105 \text{ cm s}^{-1}$ this turbulent action emptied the fluid from the hydraulic jump by forcing it over the dam and the jet became completely unstable. These observations of a transition in the flow configuration and the appearance of the periodic wave have occurred in similar experiments performed using apparatus with a double-sided orifice with or without the dam, roof or deflecting wire. In double-sided orifices, the two sides undergo the transition at slightly different speeds.

In the experiment described above, the velocity of the jet and heat current in the stream were coupled through the thermomechanical effect. In the experiments using a gravitational head, the heat current in the stream and the velocity of the jet could be independently varied. The two flow states described above occupy distinct regions on a plot of velocity versus heater power. The geometry of the orifice on the gravity apparatus was similar to that illustrated in Fig. 1, but without a dam, roof or deflecting

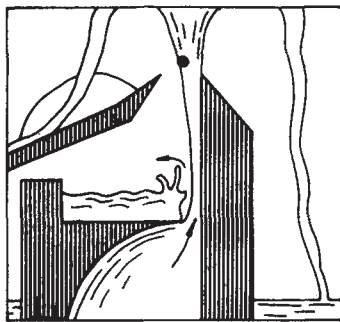


Fig. 3 A schematic drawing of the appearance of the waves on the surface of the hydraulic jump at a stream velocity of $\sim 105 \text{ cm s}^{-1}$. This has been recorded well on movie film but is not easily captured with still photography.

wire. A separate thermomechanical pump was used to fill a reservoir while it simultaneously drained through the orifice. Jet heights $\leq 7 \text{ cm}$ were achieved in this way. For jet heights $\leq 1.5 \text{ cm}$ the jet collapsed on itself otherwise the transition from the disordered to ordered flow could be induced by heating the stream (by running the heater at approximately 20 mW). (The separate heater which filled the reservoir through the thermomechanical effect had no effect on the flow near the orifice.) The values of the velocity at the orifice and the heat for which the transition in the flow was observed were found by using a constant reservoir level and increasing the current to the heater in the stream until the flow was seen to make the change to the ordered state.

Figure 4 shows a plot of velocity at the orifice against heater power for two different orifices with attack angles of 30° and 45° . The transition from the disordered to the ordered state and back again was generally observable except for low velocities where the two states become indistinguishable as the disturbances on the jet surface were reduced in amplitude as was the height of the hydraulic jump. Figure 4 shows that the ordered or low-velocity state is stabilized by heat flux. It appears that the transition would occur at zero heat flux (and somewhat lower velocity) if the observation technique was capable of resolving the difference between the two states there. In the thermomechanical effect the velocity should be proportional to the square root of the power so that the experiments using the thermomechanical effect as the driving force would lie on a locus rising in the unresolved region and crossing the transition line from disorder to order. The heat input has the effect of accentuating the velocity difference $v_n - v_s$ of the normal and superfluid components and increasing the Gorter-Mellink mutual friction in the region of the orifice⁵.

The liquid helium that flows through the small constricted stream across the tip of the region of vapour between the jet and the hydraulic jump so that the hydraulic jump may maintain itself at height h against the force of gravity, must travel at a velocity of $v = (2gh)^{1/2}$. Hence, the critical height where turbulence disturbs the periodicity of the surface waves gives a critical velocity in the narrow stream of $25 \pm 1 \text{ cm s}^{-1}$ which is very close to the observed critical velocities in saturated Rollin films of thickness $\sim 3 \times 10^{-6} \text{ cm}$. Thus the constriction would allow only the superfluid component to pass through into the hydraulic jump. Experiments done without a dam showed that a jump of about the same size could still be maintained, with a waterfall at the far edge of the pool. This implied a much larger fluid velocity in the constriction.

An explanation of the observed phenomena must contain a characteristic length which is greater than the scale of the microscopic defects on the walls of the fountain chamber, as very little effort was made to smooth the orifice and fountain vessel's walls, yet, despite the turbulent vortex tangles which these defects undoubtedly created⁶, large-scale order was observed.

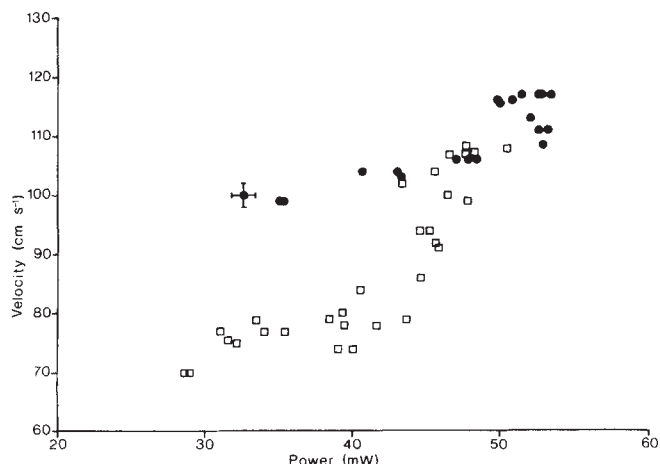


Fig. 4 The velocity of the jet and the heater power at the transition point. The closed circles are for an attack angle of 45° and the squares are for an attack angle of 30° . An example of the error associated with each point is shown with one of the points. For values of velocity and heater power above and to the left of these points, the flow was disordered. For values below and to the right the flow was ordered.

The experiment has revealed a series of detailed and unusual flow characteristics in superfluid helium not previously observed. Specifically, we observed transition to and from an ordered state with a spatial bifurcation of the jet stream. All experiments described here have been examined with high speed (3,000 frames per second) cinematography.

This work was supported by the Natural Science and Engineering Research Council of Canada.

Received 19 December 1986; accepted 21 August 1987.

1. Van Alphen, W. M., Olijhoek, J. F. & de Bruyn Ouboter, R. *Physica* **32**, 1901–1904 (1966).
2. Lorenson, C. P., Griswold, D., Nayak, V. U. & Tough, J. T. *Phys. Rev. Lett.* **55**, 1494–1497 (1985).
3. Wilks, J. *The Properties of Liquid and Solid Helium* 39–41 (Oxford University Press, London, 1967).
4. Droegemeis, K. K. & Wilhelmson, R. B. *Bull. Am. met. Soc.* **67**, 416–417 (1986).
5. Baranghi, C. F., Donnelly, R. J. & Vinen, W. F. *J. low temp. Phys.* **52**, 189–247 (1983).
6. Schwarz, K. W. *Phys. Rev. Lett.* **50**, 364–367 (1983).

Formation of organic carbon compounds from metal carbonates

A. Reller, C. Padeste & P. Hug

Institute for Inorganic Chemistry, University of Zürich, Winterthurerstrasse 190, 8057 Zürich, Switzerland

Natural metal carbonates, in particular the abundant alkaline-earth metal carbonates MgCO_3 (magnesite), $\text{MgCa}(\text{CO}_3)_2$ (dolomite) and CaCO_3 (calcite), are formed by various carbonatization processes with atmospheric carbon dioxide. They represent not only an important buffer system within the ecological carbon cycle, but also the most important feedstocks for the production of MgO (magnesia), CaO (calcia) and related chemicals. The thermal decomposition of these naturally occurring carbonates is well documented: in vacuum, inert or oxidizing atmosphere the main gaseous product is carbon dioxide, and as solid products the pure or mixed metal oxides are formed. Here we report the results of an investigation of the thermal reactivity of pure alkaline-earth metal carbonates and mixed alkaline-earth metal transition metal carbonates in a hydrogen atmosphere. We found that, compared with the analogous degradations in inert or oxidizing atmospheres, the reaction temperatures are lowered by at least 150 K . Depending

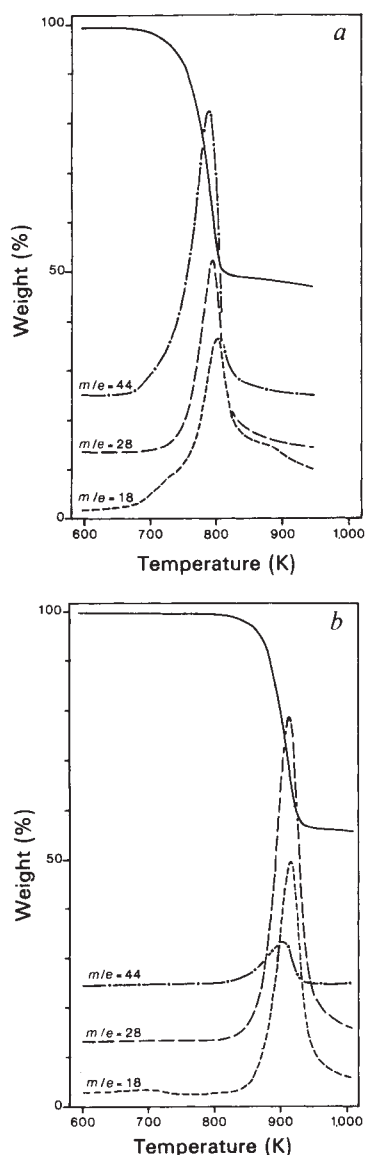


Fig. 1 Combined thermogravimetric/mass spectrometric measurements of the decompositions of the minerals magnesite (a) and calcite (b) in reducing atmosphere. Experimental conditions: sample weights were 10.99 mg (magnesite) and 8.35 mg (calcite) respectively; heating rate, 10 K min^{-1} ; atmosphere H_2 (10^5 Pa); masses detected, H_2O ($m/e = 18$), CO ($m/e = 28$) and CO_2 ($m/e = 44$). The unbroken curve is the thermogravimetric curve.

on the transition metal present in the initial mixed carbonates, CO or CH_4 is formed as the main gaseous carbon compound.

Using a combined thermomicrobalance/mass spectrometer unit, we investigated the thermal behaviour of the carbonates in a hydrogen atmosphere (10^5 Pa). In addition, the gases evolved during the processes studied were analysed by gas chromatography. As samples, magnesite from the Ural (Soviet Union), dolomite from the Binn valley (Switzerland), and calcite from the Gonzen mine (Switzerland) were used. Figure 1 shows the thermogravimetric/mass spectrometric measurements of the decomposition of magnesite and calcite respectively. The experimental details are given in the legend to Fig. 1. Magnesite decomposes in reducing atmosphere to MgO under evolution of even amounts of carbon dioxide and carbon monoxide as well as water. In the case of calcite, carbon monoxide is formed as the main volatile carbon compound ($\text{CO}:\text{CO}_2 \approx 10:1$). The predominant amount of CO formed cannot be explained by the water-gas shift, but rather indicates considerable reduction pro-

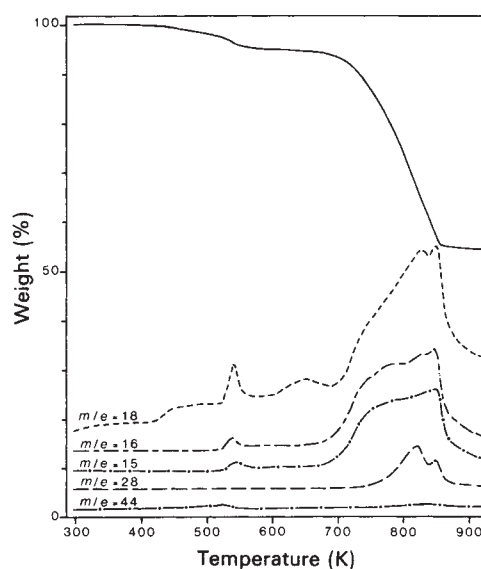


Fig. 2 Combined thermogravimetric/mass spectrometric measurement of the decomposition of calcium-cobalt carbonate ($\text{Ca}:\text{Co} = 10:1$) in reducing atmosphere. Experimental conditions: sample weight, 11.63 mg; heating rate, 10 K min^{-1} ; atmosphere, H_2 (10^5 Pa); masses detected, H_2O ($m/e = 18$), CH_4 ($m/e = 16$ and $m/e = 15$), CO ($m/e = 28$) and CO_2 ($m/e = 44$). The unbroken curve is the thermogravimetric curve.

cesses of the evolved carbon dioxide and thus a remarkable change of the mechanism of the degradations. The decomposition temperature as well as the ratios of the gaseous products for dolomite lie between the two pure carbonates. Compared with the analogous processes in non-reducing atmospheres^{1,2}, the decomposition temperatures of all three compounds are considerably lowered, by $>150 \text{ K}$. As revealed by high-resolution electron microscopy, the solid products MgO and CaO are formed as conglomerates of microcrystalline domains with diameters in the range of $10\text{--}20 \text{ nm}$. These oxides are highly reactive with respect to the formation of the respective hydroxides or to an eventual recarbonatization.

Further studies focused on the thermal reactivity of mixed alkaline-earth metal/transition metal carbonates in hydrogen atmosphere. Mg-Me- , respectively, Ca-Me- carbonates ($\text{Me} = \text{Co, Ni and Cu}$; 10% each) were prepared by co-precipitation from the respective nitrate solutions with sodium carbonate. The investigation of these degradation processes yielded remarkable results: First the decomposition temperatures of all substituted carbonates are lowered again. Compared to the decomposition temperatures of the pure carbonates in non-reducing atmospheres, a decrease in the range of 200 to, at most, 400 K (with nickel) has been observed for the degradation temperatures of the mixed carbonates in hydrogen atmosphere. As a second and rather important result we found, that depending on the transition metal present, different gaseous carbon compounds are evolved. In the case of copper, mixtures of carbon monoxide and carbon dioxide are detected. For samples doped with cobalt the decomposition yields carbon monoxide and predominantly methane (Fig. 2). The amount of carbon dioxide formed is negligible. For calcium-nickel or magnesium-nickel carbonates, the main gaseous carbon compound evolved is methane ($>90\%$). As revealed by high-resolution electron microscopy and X-ray diffractometry, the solid products consist of mixture of the microcrystalline alkaline-earth metal oxides and elemental transition metal. These products prove to be highly effective catalysts for the partial reduction of CO_2 to CO , which can be converted into hydrocarbons and/or partially oxidized hydrocarbons via syngas-reactions³, or the direct conversion of CO_2 into methane under the atmospheric pressure of hydrogen (A.R., C.P. and M. Grätzel, Swiss patent no. 2341,

1986). A detailed characterization of the respective catalytically active species is to be published.

Our experimental findings allow the following conclusions: in a hydrogen atmosphere, the thermal degradation of metal carbonates, in particular of the naturally most abundant phase calcite, is observed at comparably low temperatures. This effect is enhanced by the admixture of selected transition metal carbonates. The evolved carbon compounds, carbon monoxide from pure alkaline-earth carbonates or methane from doped carbonates, give evidence for a different mechanism of the thermal degradations. The fact that the solid products not only represent potential catalysts for the conversion of CO_2 into organic carbon compounds with various possible products, but also act as effective CO_2 -trapping systems, renders these systems interesting with respect to the problems arising from the increasing atmospheric CO_2 concentration²⁻⁶.

We thank Professor Oswald (director of the institute) for generous support as well as Professors Oberholzer (ETH Zürich) and Stalder (Naturhistorisches Museum Berne) for disposing of the mineral samples investigated.

Received 25 June; accepted 19 August 1987.

1. Brown, W. E., Dollimore, D. & Galwey, A. K. in *Comprehensive Chemical Kinetics* Vol. 22, 167-174 (eds Bamford, C. H. & Tipper, C. F. H.) (Elsevier, Amsterdam, 1980).
2. Maciejewski, M. & Reller, A. *Thermochim. Acta* **110**, 145 (1987).
3. Sachtler, W. *Proc. 8th Int. Congr. Catal.*, Vol. 1, 151-173, Berlin (Verlag Chemie, Weinheim, 1984).
4. Liss, P. S. & Crane, A. J. *Man-Made Carbon Dioxide and Climatic Change* (Geo Books, Norwich, 1983).
5. Gribbin, J., *New Scient.* **110**, 32-33 (1986).
6. Kerr, R. A. *Science* **232**, 573-574 (1986).

The stability of the fullerenes C_n , with $n = 24, 28, 32, 36, 50, 60$ and 70

H. W. Kroto

School of Chemistry and Molecular Sciences, University of Sussex, Brighton BN1 9QJ, UK

It has been proposed¹ that the geodesic and chemical properties inherent in a closed, hollow, spheroidal, carbon cage structure with the symmetry of a European football can readily explain the remarkable stability observed for the C_{60} molecule. Here I present a set of simple, empirical chemical and geodesic rules which relate the stability of carbon cages mainly to the disposition of pentagonal rings, or various directly fused pentagonal ring configurations. The rules yield cluster magic numbers consistent with observation and in particular predict that the fullerenes, C_n for which $n = 24, 28, 32, 36, 50, 60$ and 70 should have enhanced stability relative to near neighbours. These results provide further evidence for the proposal that closed hollow cages form when carbon nucleates in the vapour phase, and in particular that C_{60} buckminsterfullerene is indeed a truncated icosahedron as originally proposed¹.

It is now clear that the spectacular behaviour of C_{60} (ref. 1) observed in carbon nucleation studies using laser vaporization techniques² results from the inability of this species to take part in the general nucleation mechanism that produces large soot-like particles^{3,4}. In the original study¹ it was suggested that a truncated icosahedral cage structure (Fig. 1) provides a simple explanation. Such a molecule was likely to be aromatic and able to survive the clustering process which leads to such soot-like particles because its closed spherical form imparts geodesic stability, as well as chemical inertness owing to the absence of the reactive sites that must occur at the edge of a flat graphite sheet. It was called buckminsterfullerene because the geodesic ideas associated with the constructs of Buckminster Fuller⁵ had been instrumental in arriving at a plausible structure. It is

convenient to retain this name for C_{60} and use the name fullerene generically for the class of all closed carbon cages composed of twelve 5-membered and an unrestricted number of 6-membered rings consistent with the constructs discussed in the original patents⁵. If necessary, conventional symmetry labels can also be used; for example, buckminsterfullerene is $(I_h)C_{60}$.

It was subsequently discovered that the C_{60} molecule had been postulated independently by several other workers⁶⁻¹². In fact, an ingenious suggestion of Jones^{11,12} which discusses hollow carbon molecules in general, predates these. Theoretical work since the initial discovery (see, for example, refs 13-19) has added significantly to the weight of evidence in favour of the suggested structure. The structural assignment rests on a range of circumstantial evidence which, taken in its entirety, makes a convincing case for a closed structure.

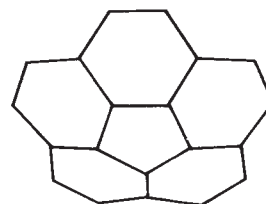
An important additional experimental observation is that, to a clear but lesser extent, enhanced stability is also shown by C_{70} and C_{50} ^{1,20,21}. Arguments that apply to C_{60} should also apply to C_{70} and C_{50} and thereby complement the case, albeit circumstantial, in favour of the closed-cage proposal¹.

A set of five basic empirical arguments can be presented to assess the relative stability of carbon cages.

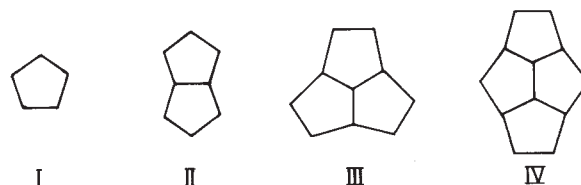
(1) The networks should conform to the usual valence requirements of carbon, in that each atom should be three-connected to other atoms by one double and two single bonds. Thus, only even-numbered cages should be stable, and many resonance structures are desirable (in C_{60} there are 12,500 resonance structures¹⁵).

(2) Polyaromatic hydrocarbons with five- and six-membered rings are abundant, but three- and four-membered rings are very unstable and seven-membered ones rare. This suggests that only 5/6-ring networks are likely to occur readily. The cages must contain 12 pentagonal rings¹², but the number of hexagonal ones is not directly restricted.

(3) From Barth and Lawton's work²² on the molecule corannulene



(seen here in an edge-on perspective to highlight its non-planar saucer shape), it is clear that a structure in which a pentagon is completely surrounded by hexagons is stable. The simple molecule in which two pentagons are fused has not so far been made, although a stabilized analogue exists. These observations suggest that a cage in which all 12 pentagons are completely surrounded by hexagons has optimum stability and one in which pentagons abut is likely to be less stable. This argument can be carried further, in that closed structures involving various fused pentagon configurations are likely to exhibit varying degrees of strain-related instability. On this simple basis it is proposed that for the following configurations



the local strain increases in order $I \rightarrow IV$. Note that this set is not complete, as multiplet configurations in which pentagons abut sequentially, that is, share only one side, are not included. Such configurations are expected to exhibit stabilities intermediate between type II and type III.

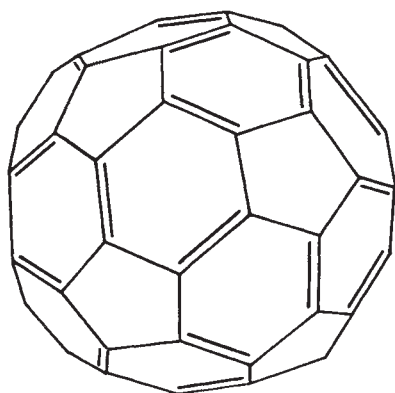


Fig. 1 The proposed structure of C_{60} buckminsterfullerene, the archetype of the fullerene family. It has I_h -icosahedral symmetry, as does the modern European football.

(4) Geodesic structural factors should favour the more symmetric isomers, which can evenly disperse the strain from bond-angle deformation. Structures in which curvature is localized or unevenly distributed are likely to be excessively strained and subject to attack at such positions.

(5) Closed-shell electronic structures are likely to be preferred. Fowler and Steer¹⁹ have shown that cages with $60+6k$ atoms (where k is 0 or ≥ 2) can have a closed-shell electronic structure. The rule is not restrictive¹⁹.

Any deductions based on such simple arguments are likely to be coloured by more subtle considerations but, leaving these aside for the moment, it is useful to see how far such simple arguments can go towards explaining the phenomena that have been observed so far.

C_{60} . C_{60} is the smallest closed network that can have all pentagons separated; that is, it involves only type-I configurations. It is thus the first highly strain-free cage capable of formation during clustering and as all 60 atoms are equivalent the strain energy is perfectly uniformly distributed. It should also have a closed-shell electronic configuration¹⁹. As C_{60} fulfills all the criteria above more perfectly than any other cluster—and, most importantly, is the first cage to form during clustering that can—it is expected to be the most stable and the dominant member of the fullerene family, as is observed.

C_{70} . It is not immediately obvious, but attempts to make model structures suggest that it is not possible to construct a closed network with separated pentagons again until the cluster has at least 70 atoms ($(D_{5h})C_{70}$, Fig. 2a). This conjecture has also been made independently by Schmalz *et al.*¹⁴. It is particularly satisfying that the rules, in their simplest form, explain both of the original observations¹: the dominance of the C_{60} and the C_{70} signals.

C_{50} . Under gentle, one-photon F_2 excimer-laser ionization, C_{50} also appears special^{20,21,23}. The necessity of having 12 pentagons in any closed cage^{11,12} indicates that in this molecule some pentagons must abut. Again attempts to make model structures lead to the conclusion that C_{50} is the smallest cluster for which a cage can be constructed without triplets of directly (type-III) or sequentially fused pentagons; that is, type-I and II configurations suffice. Schmalz *et al.* have confirmed this conjecture using combinatoric relations¹⁴. One such 50-atom cluster, $(D_{5h})C_{50}$, is shown in Fig. 2b.

C_{36} , C_{32} , C_{28} , C_{24} and C_{20} . It would appear²⁴ that all clusters with ≥ 36 atoms photo-fragment by eliminating two atoms at a time to produce the next-smaller even cluster. C_{34} appears to fragment by the loss of both two and three atoms at a time. C_{32} and smaller clusters seem to fragment almost completely into a range of small carbon fragments. The results of recent mass-spectrometric studies on carbon ions produced in hydrocarbon flames²⁵ are very interesting, in that not only are prominent

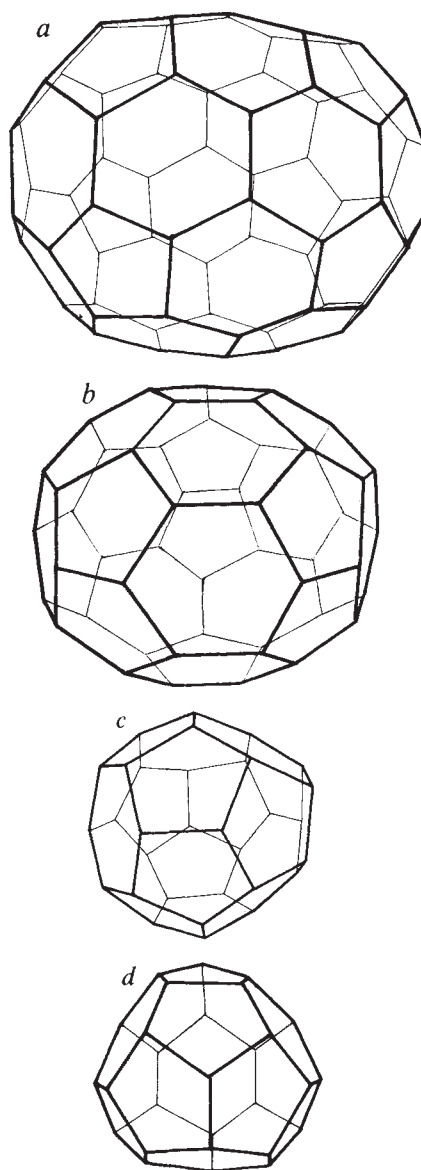


Fig. 2 Structures of various fullerenes. *a*, The most stable C_{70} fullerene, formed by separating two halves of C_{60} by a ring of ten extra carbon atoms to produce, rather appropriately, a structure similar to that of an American football. *b*, C_{50} fullerene, containing only isolated singlet and doublet pentagon configurations. *c*, A handed C_{32} fullerene viewed approximately along its threefold axis. *d*, The tetrahedral fullerene C_{28} , which bears a striking family relationship with Gombert's²⁷ organic stable free radical, triphenylmethyl, $\cdot C(C_6H_5)_3$, the first organic free radical identified and the forerunner of free-radical chemistry.

peaks at 50, 60 and 70 observed but there appear to be no peaks lower than 32. These results are clear evidence for the validity of the proposed soot formation mechanism involving open, curved, graphitic shell embryos (refs 4, 3, 26; H.W.K. and K. McKay, to be published). They also suggest that hot clusters with less than 32 atoms may be particularly unstable under these conditions. Cages with less than 36 atoms appear to require the inclusion of some type-III or higher-order configurations in the network, and it seems reasonable to suppose that the discontinuity in the photo-fragmentation behaviour that occurs between C_{36} and C_{32} is a consequence of these structural limitations. The rules suggest that the most stable 32-atom cluster should be the rather interesting fullerene, $(C_3)C_{32}$ (Fig. 2c), which involves only two type-III configurations and, incidentally, is handed. There does not appear to be a 30-atom cage

that can avoid the inclusion of high-strain type-IV configurations, and it is thus quite reasonable to expect that multi-photon fragmentation of C_{32} , which must introduce large amounts of vibrational excitation into the cages, should show a discontinuity of the kind observed. There is, however, one even smaller cluster that can avoid type-IV configurations: the tetrahedral cluster (T_d) C_{28} shown in Fig. 2d. This intriguing molecule consists of four linked, symmetrically disposed type-III configurations and four hexagonal faces. Its stability will depend on the ability of the spare electrons on the four carbon atoms at the centres of the four type-III configurations to stabilize by conjugation. C_{28} is not expected to be a closed-shell molecule (P. A. Fowler, personal communication). But it is interesting and perhaps significant, as far as stability is concerned, to note that this species is a rather neat condensed pseudo-tetramer of Gombert's (J. R. Heath *et al.*, unpublished results) famous stable organic radical, triphenylmethyl, $\cdot C(C_6H_5)_3$, which was the first free radical to be positively identified, and was the species that initiated free-radical chemistry. It is highly likely that the methane-like analogue, $C_{28}H_4$, is stable. Note that there are clustering conditions that show the C_{28} cluster peak to be dominant when F_2 excimer radiation is used to ionize the cluster distribution²⁷. Cox *et al.*²³ have also observed extra stability for C_{28} . At first sight it would seem that the out-of-plane distortion is too great and the cage network must have been deformed beyond breaking point to close such a structure involving sp^2 -hybridized carbon. Carbon, however, appears able to break so many rules that one must continually reconsider them for this element. It is, after all, only after a great deal of study that the structures of even the smallest molecules such as C_2 and C_3 have been characterized and understood theoretically. Such studies urge caution in being too hasty with regard to assumptions about the electronic and geometric structure of carbon compounds in terms of conventional valence theory. Thus, one must seriously consider the possibility that the observed special behaviour of the C_{28} signal reflects the formation of the molecule shown in Fig. 2d, even though conventional wisdom would suggest that such a cage would be too strained to exist. The only other reasonable suggestion is that the cluster is linear; however, the evidence on the shorter linear chains^{28,29} suggests that their magic-number sequence is not sensitive to the clustering conditions in the same way that the large even ones are. Finally, note that a closed C_{22} cage cannot be made¹⁹, and that the most strained molecule must be C_{20} , which as it consists only of pentagons is the smallest fullerene that could exist. The large-even-cluster mass spectrum pattern of ref. 23 shows a clear break at C_{24} , with no evidence for C_{22} or C_{20} . It seems that under cool clustering conditions and very gentle ionization C_{24} is the smallest species that can be stable, whereas under high-temperature clustering conditions or high photon irradiation C_{32} is the smallest.

The large even clusters show stability across the whole range relative to the odd ones⁴, suggesting that closure may be a general phenomenon. However, even the highly symmetric fullerenes are not expected to have closed electronic shells (P. A. Fowler, personal communication), although it is the case for C_{60} and C_{70} . Thus, just how significant the closed-shell requirement is remains an open question; in fact, the stability of Gombert's triphenylmethyl suggests that it might not be as important as has been expected for graphitic networks. In addition, the fact that C_{60} shows special behaviour whether it is neutral or positively or negatively charged implies that the closed-shell requirement may not be a major criterion of stability in general. If this conclusion is correct then it may be because these systems are ideal delocalization networks for free electrons.

The approach described here can probably be developed further to include other possible fused pentagonal configurations and mixtures thereof. But more quantitative results would probably require a much more sophisticated approach, involving not only more detailed consideration of the bond-angle strain

involved in closing up such networks but also other, more complicated problems such as electronic structure considerations. Some of these factors have been considered by various groups¹⁴⁻¹⁹. Nevertheless, the very simple arguments presented here appear to predict a special nature for C_{24} , C_{28} , C_{32} , C_{36} , C_{50} , C_{60} and C_{70} —species for which varying degrees of special behaviour have been observed. This is not an obvious magic number sequence and it would be very surprising if an alternative model that does not involve closed cages were able to arrive at the same set.

I thank David Walton and Kappa Cornforth for helpful discussions; Tom Schmalz and Patrick Fowler for advice on some of the ideas discussed here and for providing results before publication; Jim Heath and Sean O'Brien for discussions of their experimental results; Alex Nickon for discussions concerning the name; and Bob Curl and Rick Smalley.

Received 11 June; accepted 30 July 1987.

1. Kroto, H. W., Heath, J. R., O'Brien, S. C., Curl, R. F. & Smalley, R. E. *Nature* **318**, 162–163 (1985).
2. Hopkins, J. B., Langridge-Smith, P. R. R., Morse, M. D. & Smalley, R. E. *J. chem. Phys.* **78**, 1627–1637 (1983).
3. Heath, J. R., O'Brien, S. C., Curl, R. F., Kroto, H. W. & Smalley, R. E. *Accs chem. Res.* (submitted).
4. Zhang, Q. L. *et al.* *J. phys. Chem.* **90**, 525–528 (1986).
5. Fuller, R. B. *Inventions—The Patented Works of Buckminster Fuller* (St Martin's, New York, 1983).
6. Osawa, E. *Kagaku* **25**, 854 (1970).
7. Yoshida, Z. & Osawa, E. *Aromaticity* (in Japanese) (Kagakudojin, Kyoto, 1971).
8. Bochvar, D. A. & Gal'pern, E. G. *Dokl. Akad. Nauk SSSR* **209**, 610–612 (trans. 239–241) (1973).
9. Davidson, R. A. *Theor. chim. Acta* **58**, 193–231 (1981).
10. Haymet, A. D. J. *J. Am. chem. Soc.* **108**, 319–321 (1986).
11. Jones, D. E. H. *New Scientist* **245**, 3 Nov. (1966).
12. Jones, D. E. H. *The Inventions of Daedalus* 118–119 (Freeman, Oxford, 1982).
13. Disch, R. L. & Schulman, J. M. *Chem. Phys. Lett.* **125**, 465–466 (1986).
14. Schmalz, T. G., Seitz, W. A., Klein, D. J. & Hite, G. E. *J. Am. chem. Soc.* (submitted).
15. Klein, D. J., Schmalz, T. G., Hite, G. E. & Seitz, W. A. *J. Am. chem. Soc.* **108**, 1301–1302 (1986).
16. Schmalz, T. G., Seitz, W. A., Klein, D. J. & Hite, G. E. *Chem. Phys. Lett.* **130**, 203 (1986).
17. Fowler, P. W. & Woolrich, J. *Chem. Phys. Lett.* **127**, 78–83 (1986).
18. Fowler, P. W. *Chem. Phys. Lett.* **131**, 444–450 (1986).
19. Fowler, P. W. & Steer, J. I. (in preparation).
20. Liu, Y. *et al.* *Chem. Phys. Lett.* **126**, 215–217 (1986).
21. O'Brien, S. C., Heath, J. R., Kroto, H. W., Curl, R. F. & Smalley, R. E. *Chem. Phys. Lett.* **132**, 99–102 (1986).
22. Barth, W. E. & Lawton, R. G. *J. Am. chem. Soc.* **93**, 1730–1745 (1971).
23. Cox, D. M., Trevor, D. J., Reichmann, K. C. & Kaldor, A. *J. Am. chem. Soc.* **108**, 2457–2458 (1986).
24. O'Brien, S. C., Heath, J. R., Curl, R. F. & Smalley, R. E. *J. Am. chem. Soc.* (submitted).
25. Gerhardt, P., Löffler, S. & Homann, K. H. *Chem. Phys. Lett.* **137**, 306–310 (1987).
26. Kroto, H. W. *Proc. Roy. Soc.* (in the press).
27. Fieser, L. F. & Fieser, M. *Organic Chemistry* 551–555 (Reinhold, New York, 1956).
28. Heath, J. R. *et al.* *J. Am. chem. Soc.* **109**, 359–363 (1987).
29. Kroto, H. W. *et al.* *Astrophys. J.* **314**, 352–355 (1987).

Catalytic C–H bond activation on silicon dioxide overlayers

Alan B. McEwen*, Wilhelm F. Maier†‡, Ronald H. Fleming‡ & Scott M. Baumann‡

* Department of Chemistry, University of California, Berkeley, California 94720, USA

‡ Charles Evans & Associates, Redwood City, California 94063, USA

The catalytic activation of hydrocarbon C–H bonds is a fundamental process in heterogeneous catalysis. Although C–H bonds are readily activated on certain metal surfaces, the nature of the catalytic sites is still not understood. Catalytic C–H bond activation with group VIII transition metals^{1–7} is commonly studied by hydrogen/deuterium (H/D) exchange reactions of saturated hydrocarbons with D_2 . These reactions have shown that the interconversion of Π -allyl intermediates is the most facile process leading to polydeuteration on platinum and rhodium^{2–5}. Planar surfaces are the most active for C–H activation^{4,5}. Carbonaceous

† To whom correspondence should be addressed.

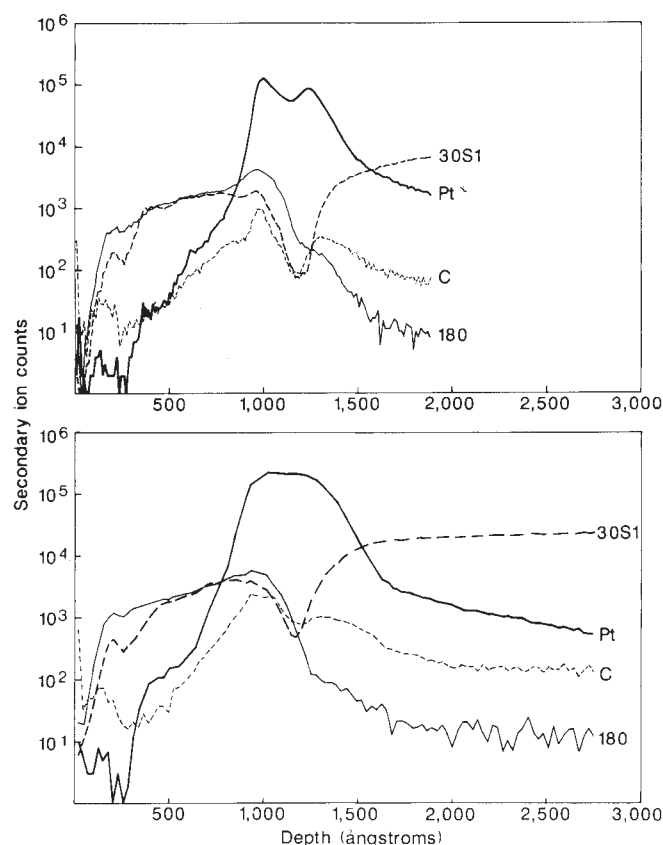


Fig. 1 SIMS depth profiles of the new and used 560 Å SiO₂/190 Å Pt/Si thin film catalyst.

surface species on platinum⁶, palladium⁷ and rhodium⁵ are not part of the active site nor do they interfere with hydrogenation or C-H activation reactions. Unfortunately, none of these numerous studies could reveal the nature of the active site or the mechanism of the initial C-H activation step. Here we present evidence that the surface of inert silica becomes catalytically active in the presence of a transition metal underlayer. The similar selectivity observed on such drastically different surfaces as silica and platinum indicates that the existence of a surface is more important than its chemical nature. These results provide new insight into the nature of C-H activation and lead the way to a novel class of well-defined catalysts based on transition metal underlayers. Our data suggest that the major action of transition metals in heterogeneous C-H activation catalysis is the activation of hydrogen and that the activated hydrogen represents the catalytically active site.

The transition metal surface in catalysis is believed to simultaneously activate C-H and H-H bonds, and to serve as a mediator for the reactive intermediates determining the product composition. The question whether radicals, charged, or covalently bonded species are involved in the mechanism has not been resolved. Many reaction mechanisms have been based on analogy with homogeneous organometallic complexes. To examine the importance of an exposed transition metal surface, transition metal film catalysts with and without a silica overlayer are compared. The rationale was that H₂ would diffuse through the overlayer to become activated at the transition metal underlayer. This activated hydrogen could then diffuse back to the gas-surface interface in a process similar to hydrogen spillover⁸. Activated hydrogen on the oxide surface was expected to exhibit characteristic catalytic properties different from those of the exposed transition metal surface. Identification of such properties would provide characteristics of hydrogen spillover catalysis.

The thin film catalysts were prepared by electron-beam

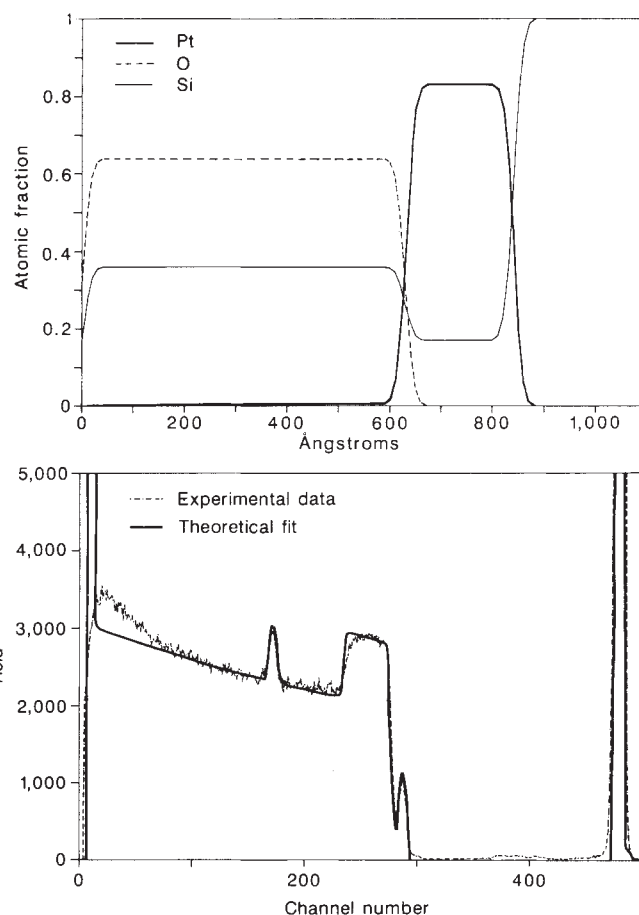


Fig. 2 RBS profile (a) and count/energy plot (b) of the used 560 Å SiO₂/190 Å Pt/Si thin film catalyst.

evaporation of platinum and SiO₂ (ref. 9) onto a polished silicon single crystal wafer to ensure uniform film thickness. Film thickness was monitored with a quartz crystal monitor. Edge effects were excluded by depositing the Pt (190 Å) through a mask leaving an outer window frame uncovered; followed by complete coverage with a thicker silica overlayer (560 Å). This silicon dioxide film catalyst is active for hexane H/D exchange at 330 °C with a turnover frequency (N_f) of 5.7×10^{13} molecules cm⁻² s⁻¹. Secondary ion mass spectrometry (SIMS)¹⁰ was used to analyse the new and used thin film catalysts (Fig. 1). Depth profiles were obtained for C, O, Si and Pt using negative ion spectrometry and Cs-ion sputtering. These analytical conditions were chosen to provide good sensitivity for C and O. Extensive charging during the analysis, preventing SIMS data acquisition, was overcome by gold coating the sample. This charging indicates that the insulating SiO₂ overlayer must be a continuous film. If cracks or holes in the overlayer were responsible for the catalytic activity platinum would be detectable throughout the depth profile (especially since an area 50 μm in diameter was analysed). In addition, an increase in carbon in the SiO₂ overlayer or at the SiO₂-Pt interface would be expected in the used film. Control experiments showed that monolayers of carbon at the SiO₂-Pt interface are readily detectable with SIMS. No evidence for platinum diffusion into the oxide overlayer was obtained.

Rutherford backscattering spectrometry (RBS)^{11,12} is more sensitive to layer thickness than SIMS. Figure 2 shows the RBS profile and the count/energy plot of the 560-Å SiO₂ overlayer film after all H/D exchange experiments we report were completed. These data were collected at grazing angle for optimal depth resolution of the thin films. The profile of this extensively used film is indistinguishable from that of the film after a single reaction at 330 °C. In the new and used films there is no trace of platinum detectable at the surface nor in the first 100 Å of

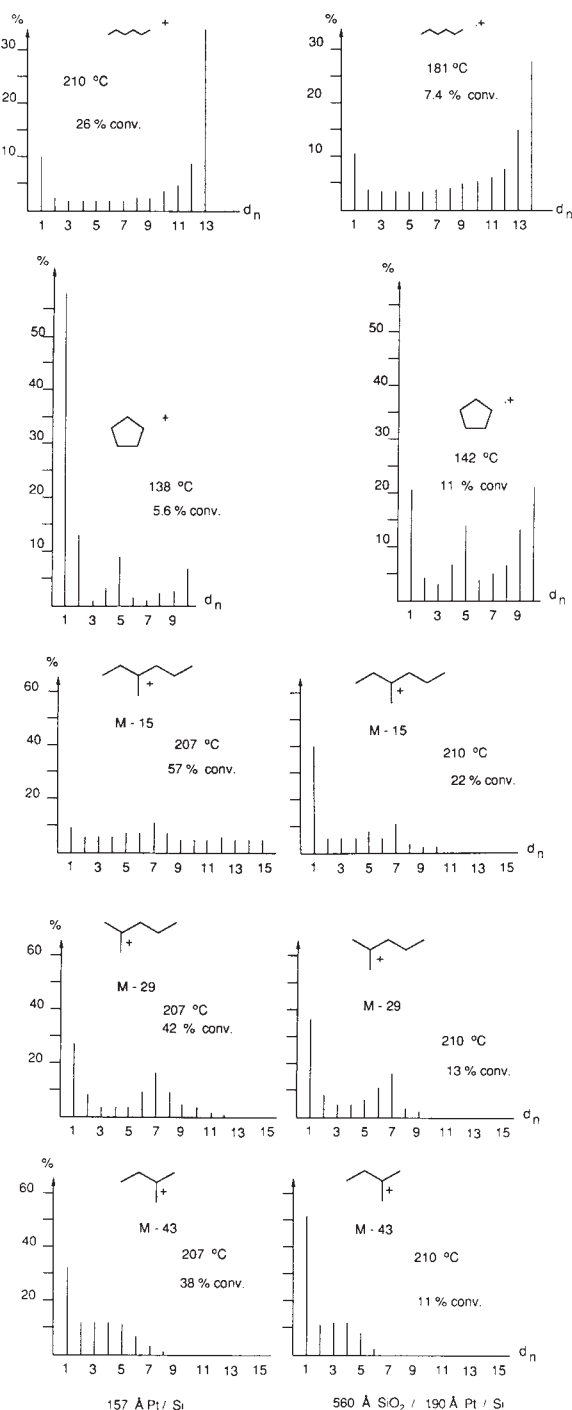


Fig. 3. Comparison of the isotope product distributions obtained from a 157 Å Pt/Si and a 560 Å SiO₂/190 Å Pt/Si thin film catalyst.

the silica. The overlayer thickness of 630 Å and the platinum film thickness of 210 Å as determined by RBS are in good agreement with the thicknesses measured during film preparation.

We have never detected any transition metals on the silica surface of similar supported films with Auger electron spectroscopy (AES). AES depth profiles have also shown the silica overlayer to be clean. The rather low sensitivity of AES (about 0.5 at.% for Pt) unfortunately reduces the usefulness of this analytical tool in this investigation. In addition, similar catalytic activity on overlayers produced by single or sequential deposition excludes activity due to cracks. Silicon dioxide deposited directly on to silicon supports is inactive.

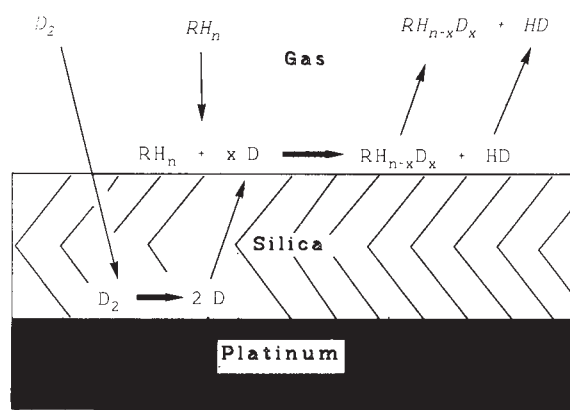


Fig. 4. A representation of the overall process of H/D exchange on silica overlayer thin film catalysts.

To examine the possibility that our observed activity may be due to platinum on the silica surface at concentrations below the detection limit of our analytical techniques ($\approx 10^{10}$ Pt atoms cm⁻²) we performed control experiments with surface platinum at concentrations from 10^8 to 10^{14} Pt atoms cm⁻² (deposited on SiO₂/Si films by impregnation techniques). No activity was observed. Since such platinum concentrations were readily detected by our analytical techniques, we conclude that there is no evidence for catalytic activity due to surface platinum.

The catalytic activity (N_t) for H/D exchange depends on silicon dioxide overlayer thickness. Exponential decay of the catalytic activity with increasing silica thickness is observed. After approximately 500 Å of silicon dioxide this activity levels at a constant value (6×10^{13} molecules cm⁻² s⁻¹) which is not affected by further thickness increase. In related hydrogenation, dehydrogenation, and H₂/D₂ exchange studies, catalytic activity shows a similar dependence on the overlayer thickness. This decrease of catalytic activity is attributed to irreversible recombination during the diffusion of the dissociated hydrogen through the overlayer. Although the overall decrease in activity by a factor of ~ 70 from the uncovered platinum film is significant, a similar drop in N_t for C-H activation is observed on platinum or rhodium catalysts by simply changing the catalyst dispersion^{4,5}. Activity decrease by annealing at elevated temperatures indicates that the catalytic activity may be based on defects in the amorphous silica rather than simple transport through an ordered silica matrix.

At single-surface interaction conditions isotopic product distributions provide a characteristic fingerprint of the catalyst's selectivity properties. The corrected isotopic product distributions in several hydrocarbons after H/D exchange on this silicon dioxide film catalyst are compared with the distributions obtained using the platinum film catalyst (Fig. 3). The isotopic distribution of *n*-hexane obtained with the silicon dioxide film catalyst is virtually identical to that obtained on the exposed platinum and other transition metal catalysts¹⁻⁵. Formation of both mono- and perdeuterated products at low conversion on the oxide overlayer surface is especially remarkable. To further probe the similarities in the isotopic distribution on the two surfaces the reaction was studied with cyclopentane and 3,3-dimethylhexane. Cyclopentane has been found to exchange exclusively on one side of the ring with heterogeneous catalysts leading to a D₅ maximum^{13,14}. This D₅ maximum is clearly present on both the platinum film and the silica overlayer. Perdeuteration in 3,3-dimethylhexane is limited to the propyl side chain leading to a D₇ maximum, a H/D exchange selectivity characteristic for quaternary hydrocarbons^{1,15}. The D₇ maximum is attributed to interconverting π -allyl intermediates, while complete perdeuteration is prevented by the quaternary carbon^{4,5}. This exchange selectivity is also observed on the silicon dioxide film catalyst. The M-15, M-29 and M-43 fragmentation

patterns confirm that polydeuteration is limited to the propyl group.

The activation energies (E_a) for hexane and cyclopentane H/D exchange (obtained from Arrhenius plots between 100 °C and 270 °C) on the 560 Å SiO₂/190 Å Pt/Si catalyst are identical (within experimental error) to those obtained on the 157 Å Pt/Si thin film [hexane (9 kcal mol⁻¹) and cyclopentane (4 kcal mol⁻¹)]. At temperatures from 270 °C to 350 °C non-Arrhenius behaviour is observed on both films. This indicates that the rate determining step of this reaction on the silica and on the platinum surface are identical.

Our results suggest that inert silicon dioxide is an active catalyst for alkane C-H activation, provided a transition metal underlayer is present (Fig. 4). In order to account for the similar catalytic behavior observed on such drastically different surfaces as silica and platinum, we postulate that activated H₂ (dissociated hydrogen) represents the active site on these catalysts and that exposed transition metal is not required. Our results suggest that H/D exchange patterns are not characteristic of bonding interactions between the hydrocarbon substrate and transition metal, and the most important action of transition metals in C-H activation catalysts is H₂ activation.

The importance of dissociated hydrogen on solids in heterogeneous catalysis has been suggested already in 1912 by Langmuir¹⁶. Related phenomena are catalysis on carbonaceous and other overlayers^{5-7,17}, enantioselective hydrogenation with modified Ra-Ni¹⁸, hydrogenation and dehydrogenation on gold surfaces^{19,20} and spillover catalysis in general.⁸ The importance of these results for other hydrogen based, heterogeneously catalysed reactions is being examined. A noteworthy selectivity

difference between the catalysis on the silica surface and on the exposed metal surface was observed with the hydrogenation of 1,2-dimethylcyclohexene and 2-hexyne which will be reported elsewhere. New catalysts based on metal underlayers with novel overlayer materials may be developed which could provide unique selectivity, activity, and stability against poisoning.

This work was supported by the National Science Foundation. The use of equipment provided by the Director, Office of Basic Energy Research, Office of Basic Energy Sciences, Materials Science Division of the US Department of Energy is gratefully acknowledged.

Received 16 March; accepted 28 July 1987.

1. Anderson, J. R. & Kemball, C. *Proc. R. Soc.* **223**, 361-377 (1954).
2. Rooney, J. J. *J. Catal.* **2**, 53-57 (1963).
3. Clarke, J. K. A. & Rooney, J. J. *Adv. Catal.* **25**, 125-183 (1976).
4. Lebrilla, C. B. & Maier, W. F. *J. Am. chem. Soc.* **108**, 1606-1616 (1986).
5. Cogen, J. M. & Maier, W. F. *J. Am. chem. Soc.* **108**, 7752-7762 (1986).
6. Inoue, Y. *et al. J. Catal.* **53**, 401-413 (1978).
7. Beebe, T. B. Jr & Yates, J. T. Jr *J. Am. chem. Soc.* **108**, 663-671 (1986).
8. Conner, W. C. Jr, Pajonk, G. M. & Teichner, S. J. *Adv. Catal.* **34**, 1-79 (1986).
9. Auwärter, M. *New Synth. Meth.* **3**, 43-58 (1975).
10. Heiland, W. & Taglauer, E. *Meth. exp. Phys.* **22**, 299-348 (1985).
11. Evans, C. A. Jr & Blattner, R. J. *A. Rev. Mater. Sci.* **8**, 181-212 (1978).
12. Chu, W. K., Mayer, J. W. & Nicolet, M. A. *Backscattering Spectrometry* (Academic, New York, 1978).
13. Anderson, J. R. & Kemball, C. *Proc. R. Soc.* **A226**, 472-489 (1954).
14. Burwell, R. L. Jr, Shim, B. K. & Rowlinson, H. C. *J. Am. chem. Soc.* **79**, 5142-5148 (1957).
15. Burwell, R. L. Jr *Accts chem. Res.* **2**, 289-296 (1969).
16. Langmuir, I. *J. Am. chem. Soc.* **34**, 1310-1325 (1912).
17. Sachtler, I. W. A., Van Hove, M. A., Biberan, J. P. & Somorjai, G. A. *Phys. Rev. Lett.* **45**, 1601 (1980).
18. Izumi, Y. *Adv. Catal.* **32**, 215-271 (1983).
19. Wood, B. J. & Wise, H. J. *Catal.* **5**, 135-145 (1966).
20. Yolles, R. S., Wood, B. J. & Wise, H. J. *Catal.* **21**, 66-69 (1971).

Evidence from en-echelon cross-grain ridges for tensional cracks in the Pacific plate

E. L. Winterer* & D. T. Sandwell†

* Scripps Institution of Oceanography, La Jolla, California 92093, USA

† Institute for Geophysics, University of Texas, Austin, Texas 78712, USA

Sea-floor topography in the Pacific is mainly aligned with original spreading directions¹, but is overprinted by alignments created by mid-plate processes. Spreading produces abyssal hills and fracture zones, and mid-plate volcanism generates seamounts, isolated or in chains. A different category of topography, the 'Cross-grain', discovered in geoid-height data collected by the Seasat radar altimeter², comprises linear troughs and swells spaced ~200 km apart, oblique to fracture zones and abyssal hills but parallel to the Hawaiian chain. Three models have been proposed for the Cross-grain: small-scale convection, organized into longitudinal rolls by the shear of the Pacific Plate²; compressive buckling³; and lithospheric boudinage resulting from plate-wide tensile stresses^{4,5}. None of the previously available data ruled out any of these models. Here we report multi-beam bathymetric data revealing long, narrow en-echelon ridges along the Cross-grain, interpreted as evidence of tension cracks in the Pacific plate.

In the equatorial Pacific, cross-grain lineations extend for more than 4,000 km, crossing from young (5 Myr) sea floor near the East Pacific Rise to Cretaceous (100 Myr) sea floor near the Line Islands. The topographic relief associated with the cross-grain geoid undulations is typically only ~200 m and, although difficult to detect in available bathymetric profiles, has been confirmed by shipboard measurements of gravity and bathymetry on crust of age 6-35 Myr (ref. 6). Another manifestation of the cross-grain pattern appears in the Line Islands region as

Cross Trend⁷ ridges trending ESE, oblique to the main SSE trend of the Line Islands.

To study the Cross-grain, we explored an area in the equatorial Pacific (see Fig. 1) from RV *Thomas Washington* in February and March 1987. This region is especially suitable for study of the Cross-grain because it shows strongly here on the Seasat maps and profiles, and appears to continue westward as the Line Islands Cross Trend. The region is blanketed by a thick (300-1,000 m) cover of pelagic sediments, with basal ages from late Cretaceous at the west end of the area to Eocene at the east end, and with an internal stratigraphy to constrain the age of formation of the Cross-grain. Neogene seismic stratigraphy in the region is tied through drill holes to the biostratigraphic timescale⁸. Strong, discontinuous amplitude anomalies in the seismic reflection records^{9,10} are spatially associated with one of the prominent cross-grain highs and resemble the seismic signature of mid-plate volcanic sills and flows in Deep Sea Drilling Project holes^{11,12}.

Bathymetry was surveyed using the Seabeam multi-beam sonar¹³. Seismic reflection profiling, using an 80-in³ water-gun source, digital recording and processing to deconvolve the source signal, was used to detect volcanic units within the pelagic sediments, to fix their stratigraphic level⁸ and to map the basement relief. To confirm and refine the Seasat observations, gravity was measured with a Bell BGM-3 gravimeter. Navigation was from transit and available Global Positioning Service satellite fixes.

Data collected along a zig-zag pattern during the first part of the expedition (Fig. 1) confirmed the existence of the cross-grain swells seen in the Seasat data. Beneath the sediments, broad (~200 km) undulations in the basement topography correlate with highs and lows in the Seasat-derived gravity², and with our shipboard gravity measurements. Exploration along the axis of the most prominent cross-grain high in the area, which extends south-east from Christmas Island (Fig. 1), revealed an unexpected series of narrow (7-15 km) ridges, en-echelon to the trend of the swell, which could be traced for ~1,000 km.

Individual ridges range in length from ~50 to 400 km and

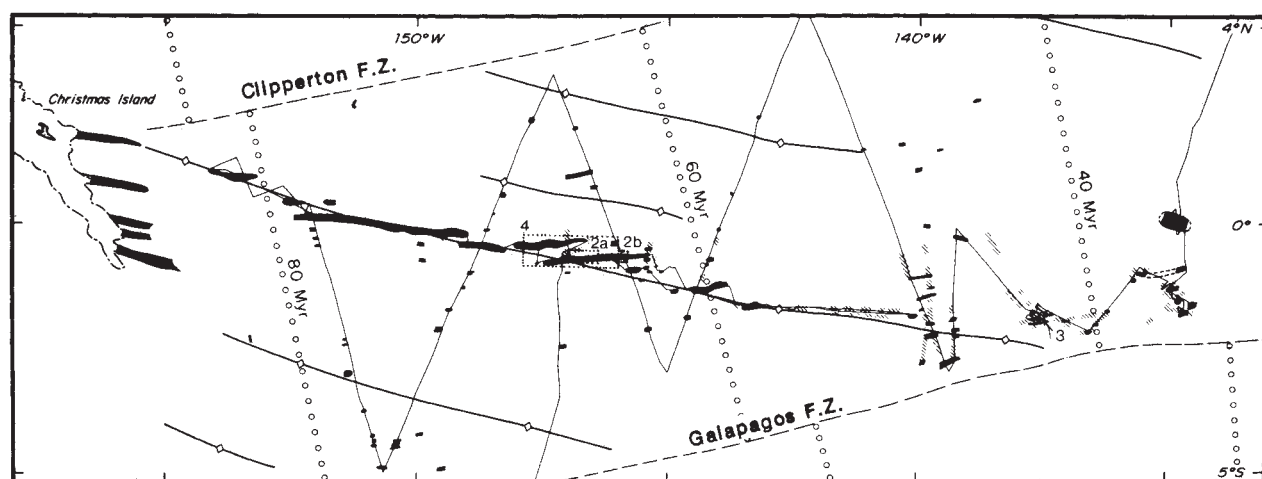


Fig. 1 En-echelon ridges, seamounts and volcanic sills and flows in the equatorial Pacific, where known from bathymetric and seismic reflection survey lines. Ridges and seamounts are shown in solid black and sills/flows by diagonal ruling. Solid lines with diamond symbols, axes of cross-grain swells seen on Seasat geoidal-height map²; dashed lines, fracture zones; dotted line, 4,000-m isobath on Line Islands ridge; thin line, track of RV *Thomas Washington*; open-circle lines, approximate positions of crustal isochrons. Locations of Figs 2-4 are shown by dotted rectangles (Figs 2, 4) and arrow (Fig. 3).

trend 10–20° more easterly than the summit line of the cross-grain swell. The en-echelon pattern steps southward in increments of ~35 km to stay mainly within ~100 km of the summit line. The known length of overlap of adjacent ridges is at least 90 km, but our survey is incomplete on this point. The ridges are continuous but not perfectly straight; short segments deviate from the main direction (095°) to follow the trend of fracture zones (077°).

The relief of the ridge system gradually decreases eastward. Beginning in the Line Islands, the ridges are high (2–3 km) and massive. The shallowest ridges form the Line Island Cross Trend⁷. Between 154 and 145° W, the ridges are generally ~1–2 km high and continuous, and east of 145° W they decrease further in height, length and continuity. Detailed surveys of two small areas at 136 and 138° W revealed only short ridges or chains of small, separated seamounts.

The ridges change morphology with size. The tallest ridges are 10–15 km wide and rise 2–3 km above the basement. They have steep flanks and relatively flat summit areas from 1 to 2 km wide. Intermediate-size ridges (1–2 km high) are narrower (7–10 km) and are topped by volcanic cones (see Fig. 2). The cones are mainly spaced at intervals of 3–7 km along the ridge, and in some places alternate from one side of the ridge to the other, like large stepping stones. The smallest ridges comprise volcanoes connected beneath the sediment cover. At wide intervals, large seamounts occur along even the smaller ridges. Where ridges turn to follow the fracture zone direction, the morphology commonly changes from 'stepping stones' to a razor-back ridge.

As the ridge system decays eastward, the expression of volcanism changes from volcanic edifices to a mixture of small volcanoes and flanking sills (and perhaps flows) intercalated within the sediments between volcanoes. Several lines of evidence

support the identification of the strong reflectors as sills: (1) they are intercalated within the normal sequence of sedimentary reflectors^{9,10} (Fig. 3); (2) they cut across sedimentary layers in some places (Fig. 3); (3) they occur at many stratigraphic levels, but not above a certain regional reflector known to be ~14 Myr old⁸. The highest units may be flows rather than sills. (4) They closely resemble discontinuous strong reflectors identified elsewhere as intrusive volcanic sills by drilling^{11,12}.

Sills and/or flows strongly tend to occur close to the en-echelon ridges and lines of small seamounts, and we suppose these to be their sources. A detailed survey, near 138° W, showed sills/flows in sedimentary troughs between aligned seamounts, either level-ponded or slightly sloping. The distribution of sills (Fig. 1) is complementary to that of the ridges. Going eastward, sills first appear at about 147° W, between ridges, and increase in areal coverage farther east. They are most concentrated in a band within 100 km of the summit line of the swell, but occur as far away as 200 km.

The morphology and distribution of the en-echelon ridges suggest that they result from filling of tensional cracks in the lithosphere. We speculate that the widths of the cracks are about equal to the width of the relatively flat summits with volcanic peaks (1–2 km). The cracks are segmented such that they always lie on the cross-grain swell. This suggests that either this cross-grain swell is a broad zone of tensional deformation of the ductile lower lithosphere or the cracks develop on the weakest lithosphere, on the summit of the swell. We favour the first hypothesis because it also explains the en-echelon pattern of surface cracks. Experimental and theoretical studies of thin ductile sheets under tension show that plastic yielding occurs along diffuse flow layers (Luders' line) oriented 55–60° from the direction of tensile stress; both extensional and shear deforma-

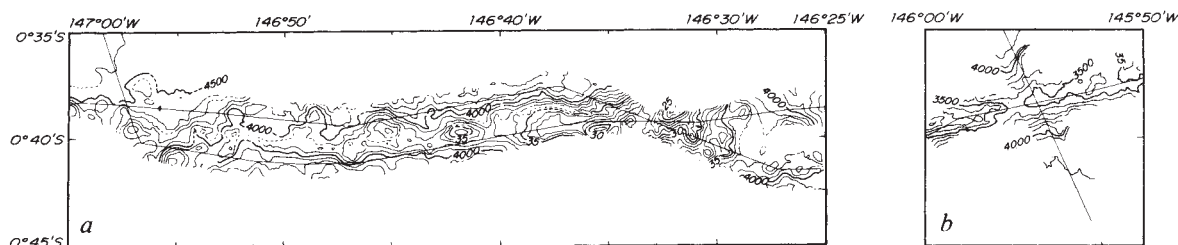


Fig. 2 Two segments of a typical en-echelon ridge, surveyed with a multi-beam swath-mapping system. The western segment (a) is mainly in the overlap zone of two adjacent ridges (Fig. 1), and descends westward from the beginning of the overlap, at 146.5° W (see Fig. 4). Summit peaks are staggered along a zone 2–3 km wide. Ship tracks are shown by thin lines; contour interval, 100 m.

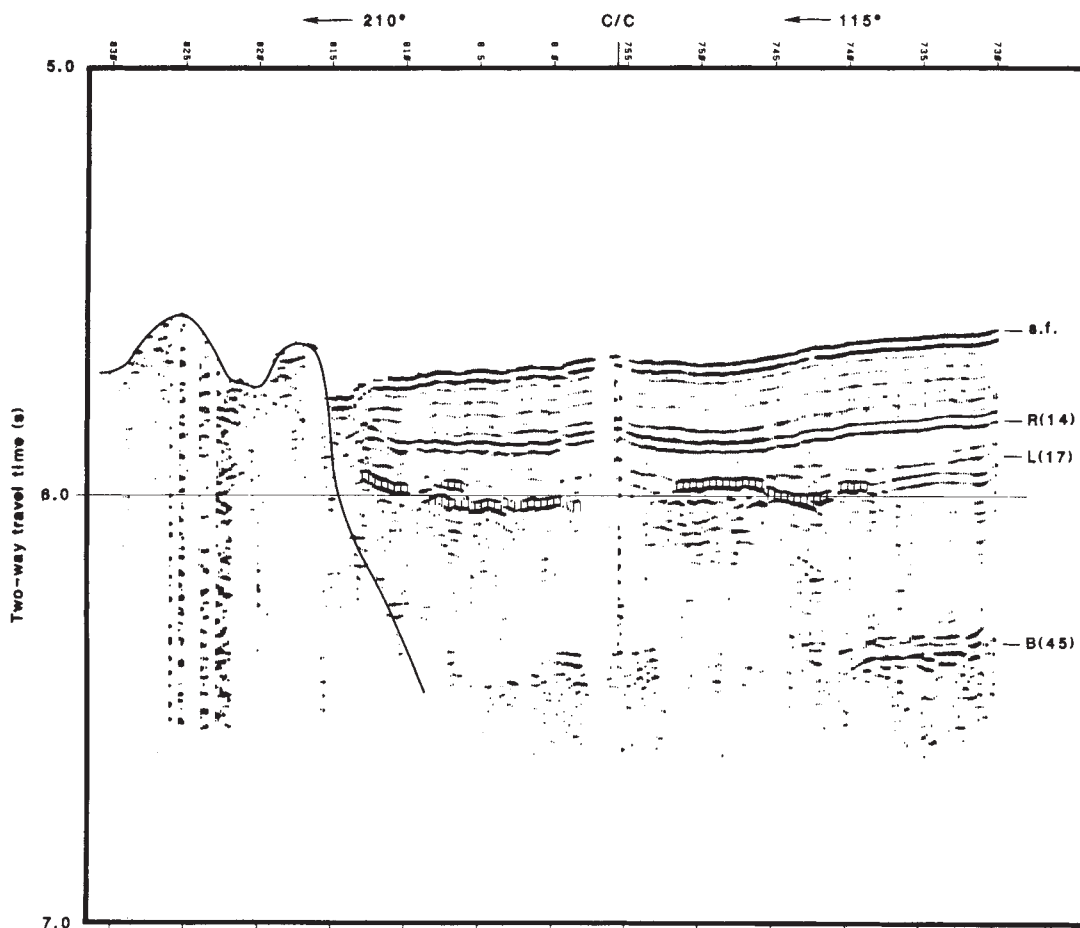


Fig. 3 Seismic reflection profile of volcanic sills intercalated in the sedimentary section. Note that the sills cut across sedimentary reflectors. The sills are identified by vertical ruling, and the volcanic ridge at left is outlined; otherwise, nothing on the original seismic record has been changed. Reflectors: s.f., sea floor; R, red reflector; L, lavan-der reflector; B, oceanic basement. Numbers in parentheses give the age of each reflector in Myr. At 1.8° S, 137.5° W (ref. 9). C/C denotes course change. 115° and 210° are the courses of the ship along the direction given by the arrows.

tion occur within the flow layers¹⁴. If the cross-grain swell is due to ductile flow in the lower lithosphere, then the en-echelon ridges are the response of the upper brittle lithosphere to extension and shear deformation: extension opens the cracks and shear produces the en-echelon pattern. If this model is correct, then the central Pacific is being subjected to a large tensional stress oriented roughly north-south. This orientation is consistent with local intraplate earthquakes, which show plate-wide tension oriented NNE, approximately perpendicular to the direction of absolute plate motion¹⁵.

Assuming that the tensional cracks are related to the formation of the Cross-grain, we can reject the lithospheric buckling model of cross-grain formation. We favour boudinage over the small-scale convection model because it requires tension and explains the opening of the cracks as well as their en-echelon pattern. On the other hand, our observations do not rule out the convection model, as small-scale convection beneath the lithosphere can also produce tensional stresses leading to cracking, as well as provide the excess temperatures required to produce melting and hence volcanism (B. Parsons, personal communication). Further work is needed to discriminate between the two models.

Our data do not indicate the source of the tensile stress, but possibilities include: (1) non-uniform cooling and contraction of the lithosphere⁵. Although most of this thermoelastic stress develops in the first 10 Myr and may be partially relieved along transform faults, this stress relief mechanism is less efficient when the spreading rate is high and transform faults are widely spaced¹⁶. (2) Plate-boundary forces such as slab pull¹⁷. (3) Drag at the base of the plate in the direction of absolute plate motion. This may produce compression in that direction and extension perpendicular to the cross-grain direction.

Given the angle between the Cross-grain and the individual ridges (10–20°), the ridge length and spacing are geometrically related. The crack spacing is relatively uniform (~35 km) and

may be controlled by the thickness of the elastic part of the lithosphere. Ridge length therefore depends only on the angle between the cross-grain swell and the crack. This explains why ridges are shorter toward the north-west, as the angle between ridges and cross-grain swell increases (see Fig. 1).

Our simple analysis of en-echelon cracks indicates that the shear stress between two cracks decreases as the overlap increases. In the one place where we surveyed the overlap of adjacent ridges, the overlap is at least 90 km, or about three times the crack spacing. The overlapping-crack model predicts that the sum of the two crack widths in the overlap region is equal to the width of the single crack in the non-overlap region. Assuming that ridge height is related to crack width, each ridge should gradually diminish as the crack width decreases in the overlap region. A plot of ridge-crest elevations (Fig. 4) confirms this prediction.

We suggest that the progression of volcanic features, from high ridges in the west to discontinuous short alignments of

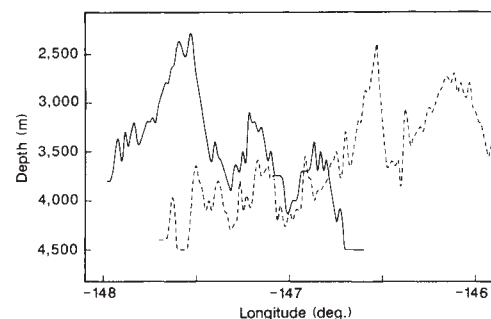


Fig. 4 Depth along the summits of two overlapping en-echelon ridges. In the overlap region, the northern ridge (solid curve) diminishes as the southern ridge grows. At depths of $\geq 4,500$ m the ridges are buried by pelagic sediments.

small seamounts and ridges, interspersed with sills in the east, reflects an eastward change of some combination of tensional crack distribution and the volume rate of magma generation: either the thermal anomaly beneath the tension cracks diminished or the cracks themselves narrowed. The two effects may be combined.

The eastward sequence may also represent a time sequence, formed progressively as the Pacific plate moved northward. We have as yet only two indications of age along the Cross-grain: middle Miocene (14 Myr) sills/flows at 138° W, and possible late Eocene (36–40 Myr) for certain Line Islands Cross Trend features^{18–20}. This age difference along the cross-grain is ~25 Myr and the distance is ~2,000 km, giving an average rate, assuming progression, of ~8 cm yr⁻¹, a figure close to the progression rate of the Hawaiian chain. Several ridges in the Line Islands Cross Trend appear to end westward in an elbow, congruent with the elbow junction between the Emperor and Hawaiian chains, also late Eocene in age²¹.

This work was supported by the US Office of Naval Research. We thank Barry Parsons for a critical review of the manuscript.

Received 14 May; accepted 12 August 1987.

1. Mammerickx, J. & Smith, S. M. *Geol. Soc. Am. Map and Chart Ser.* **26**, (1978); **43**, (1980). (1980).
2. Haxby, W. F. & Weissel, J. K. *J. geophys. Res.* **91**, 3507–3900 (1986).
3. McAdoo, D. C. & Sandwell, D. T. *J. geophys. Res.* **90**, 8563–8569 (1985).
4. Turcotte, D. L. & Oxburgh, E. R. *Nature* **244**, 337–339 (1973).
5. Ricard, Y. & Froidevaux, C. *J. geophys. Res.* **91**, 8314–8324 (1986).
6. Parsons, B., Weissel, J. K., Daly, S., Buck, R. & Haxby, W. F. *Eos* **67**, 356 (1986).
7. Winterer, E. L. *Init. Rep. DSDP* **33**, 731–747 (1976).
8. Mayer, L. A., Shipley, T. H. & Winterer, E. L. *Science* **233**, 761–764 (1986).
9. Shipley, T. H. *et al. Init. Rep. DSDP* **85**, 825–837 (1985).
10. Winterer, E. L. & Shipley, T. H. *Eos* **65**, 977 (1984).
11. Winterer, E. L. *et al. Init. Rep. DSDP* **17**, 247–262 (1973).
12. Lancelot, Y. *et al. Init. Rep. DSDP* **41**, 233–326 (1977).
13. Renard, V. & Allenou, J.-P. *Int. hydrogr. Rev.* **56**, 35–67 (1979).
14. Nadai, A. *Theory of Flow and Fracture of Solids*, 320–326 (McGraw-Hill, New York, 1950).
15. Okal, E. A., Talandier, J., Sverdrup, K. A. & Jordan, T. H. *J. geophys. Res.* **85**, 6479–6495 (1980).
16. Sandwell, D. T. *J. geophys. Res.* **91**, 6405–6417 (1986).
17. Wortel, R. & Cloetingh, S. *Geology* **9**, 425–430 (1981).
18. Schlanger, S. O. *et al. J. geophys. Res.* **89**, 11261–11272 (1984).
19. Winterer, E. L. *et al. Init. Rep. DSDP* **17**, 47–102 (1973).
20. Premoli-Silva, I. & Brusa, C. *Init. Rep. DSDP* **61**, 439–473 (1981).
21. Jackson, E. D. *et al. Init. Rep. DSDP* **55**, 5–31 (1980).

Evidence for a genetic aetiology in reading disability of twins

J. C. DeFries, David W. Fulker & Michele C. LaBuda

Institute for Behavioral Genetics, University of Colorado, Boulder, Colorado 80309-0447, USA

Reading disability (dyslexia) is a major social, educational, and mental health problem. Although estimates of prevalence vary, up to 10–15% of school-age children have severe reading deficits in spite of average intelligence and adequate educational opportunity¹. That reading disability may have a constitutional basis has long been recognized², and results of twin and family studies suggest that one or more of its forms may be heritable^{3,4}; however, definitive evidence for a genetic aetiology has not been reported. Establishing a heritable basis for reading disability could suggest possible causes, give improved risk estimates, facilitate early diagnosis, and provide validity tests for ostensible subtypes. In this report, we apply a recently developed multiple regression analysis^{5,6} to data collected from a sample of 64 pairs of identical twins and 55 pairs of fraternal twins, in which at least one member of the pair is reading disabled, and present evidence for a significant genetic aetiology.

In twin studies of pathology, a comparison of concordance rates for a particular disorder in identical (MZ) and fraternal (DZ) twins provides a test for genetic aetiology. When index

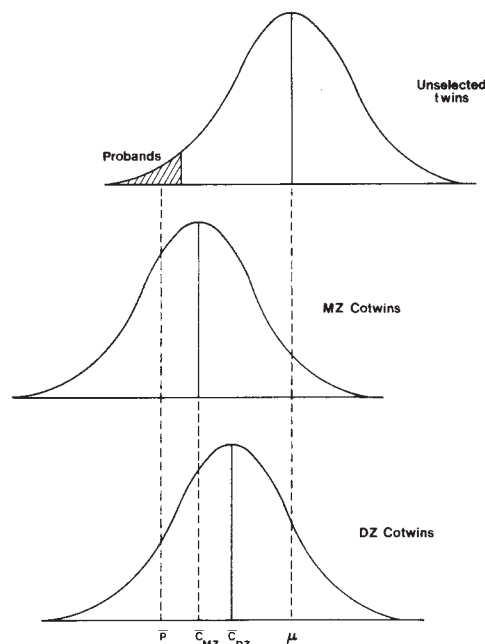


Fig. 1 Hypothetical distributions for reading performance of an unselected sample of twins, and of the identical (MZ) and fraternal (DZ) cotwins of probands with a reading disability. The differential regression of the MZ and DZ cotwin means toward the mean of the unselected population (μ) provides a test of genetic aetiology.

cases are chosen because of their extreme scores on some continuous variable (such as reading ability), a more appropriate test is to compare the means of the cotwins of the MZ and DZ probands. For example, when probands have been identified by low test scores (see Fig. 1), the scores of both MZ and DZ cotwins will regress toward the mean of the unselected population. But, to the extent that the condition is heritable, this regression toward the mean should differ for MZ and DZ cotwins. Because the coefficient of relationship for MZ twins is one, whereas that for DZ twins is one-half, scores of DZ cotwins should regress more toward the mean of the unselected population than do those of MZ cotwins. Thus, if the means for MZ and DZ probands are equal, a simple *t*-test of the difference between the means for MZ and DZ cotwins would be a test for genetic aetiology. Fitting the following multiple regression model to the twin data, however, provides a more general and statistically more powerful test:

$$C = B_1P + B_2R + A$$

where C is a cotwin's predicted score, P is the proband's score, R is the coefficient of relationship, and A is the regression constant. B_1 , the partial regression of cotwin's score on proband's score, is a measure of average twin resemblance. B_2 , the partial regression of cotwin's score on the coefficient of relationship, equals twice the difference between the means for MZ and DZ cotwins after covariance adjustment for the difference between MZ and DZ probands. Thus, the significance of B_2 provides a direct test for genetic aetiology. Moreover, the ratio of B_2 to the difference between the means for probands and for the unselected population gives a measure of the extent to which this difference is due to heritable influences⁵.

As a part of the Colorado Reading Project⁷, an extensive psychometric test battery, which includes 11 subtests from the Wechsler Intelligence Scale for Children-Revised (WISC-R, excluding Mazes) (ref. 8) or the Wechsler Adult Intelligence Scale-Revised (WAIS-R) (ref. 9), is being administered to MZ and DZ twin pairs, with at least one member of each pair reading disabled, and to a comparison group of control twins, who are normal readers. To minimize the possibility of ascertainment

Table 1 Mean discriminant scores and standardized test scores of 64 identical (MZ) and 55 fraternal (DZ) reading-disabled probands and their cotwins expressed as deviations from control means

Measure	Identical		Fraternal		B_1	B_2
	Proband	Cotwin	Proband	Cotwin		
Discriminant	-2.75	-1.98	-2.79	-1.63	$0.77 \pm 0.08^*$	$-0.77 \pm 0.27^*$
Reading Recognition	-2.30	-1.65	-2.39	-1.32	$0.67 \pm 0.08^*$	$-0.79 \pm 0.28^*$
Reading Comprehension	-1.89	-1.36	-1.76	-1.10	$0.50 \pm 0.10^*$	-0.38 ± 0.30
Spelling	-1.90	-1.29	-1.89	-1.02	$0.50 \pm 0.09^*$	$-0.53 \pm 0.30^*$
Coding	-0.64	-0.63	-0.97	-0.78	$0.61 \pm 0.09^*$	-0.12 ± 0.31
Colorado Perceptual Speed	-1.23	-1.06	-1.24	-1.04	$0.53 \pm 0.09^*$	-0.05 ± 0.27
Digit Span	-0.89	-0.71	-1.14	-0.52	$0.51 \pm 0.10^*$	$-0.62 \pm 0.33^*$

B_1 is the partial regression of cotwin's score on proband's score, a measure of average MZ and DZ twin pair resemblance. B_2 is the partial regression of cotwin's score on the coefficient of relationship and equals twice the difference between the means of MZ and DZ cotwins after covariance adjustment for the difference between probands. The sample of reading-disabled twin pairs includes 39 female MZ, 25 male MZ, 26 female DZ, and 29 male DZ. Although there are significant gender differences, with males typically scoring lower than females, the average scores of MZ and DZ probands do not differ significantly and gender differences do not differ as a function of zygosity. When an augmented regression model that included gender as a dummy variable was fitted to these data, there was no evidence for differential genetic aetiology as a function of gender.

* $P < 0.05$, one-tailed test.

bias, the twins' school records are reviewed for reading problems, such as low reading-achievement scores, referral to resource rooms because of poor reading ability, or reports by school psychologists. A discriminant function score is computed for each member of the pair based upon three measures of reading performance (Reading Recognition, Reading Comprehension, and Spelling subtests of the Peabody Individual Achievement Test¹⁰), two subtests from the WISC-R or WAIS-R (Coding-B and Digit Span), and the Colorado Perceptual Speed Test¹¹. Discriminant weights were estimated from an analysis of independent data obtained from a sample of 140 reading-disabled and 140 control non-twin children in which the rate for correct classification was ~95%. To be included in our sample, at least one member of each prospective reading-disabled twin pair must be classified as affected by the discriminant analysis, and the twin with the lower discriminant score is designated as the proband. In addition, probands must have a minimum IQ of 90 (on either the Verbal or Performance scale of the WISC-R or the WAIS-R), show no evidence of neurological, emotional, or behavioural problems, and have no uncorrected visual or auditory acuity deficits. Zygosity is determined with 95% accuracy using the Nichols and Bilbro¹² questionnaire and confirmed if necessary by blood test. Results described in the present report are based upon analyses of data from 119 twin pairs (mean age 12.7 years).

The average discriminant scores and standardized test scores of the MZ and DZ probands and their cotwins are presented in Table 1. Because probands were selected on the basis of the discriminant score, the fit of the basic regression model to data for that measure is appropriate as a test of genetic aetiology. The average discriminant scores of MZ and DZ probands are -2.75 and -2.79, respectively, over 2.5 standard deviations below the mean for matched control twins. The means for their cotwins are -1.98 and -1.63. Thus, the scores of the DZ cotwins have regressed 1.16 standard deviation units on the average toward the control mean, whereas those of the MZ cotwins have regressed only 0.77 standard deviation units. The highly significant B_2 estimate ($P = 0.003$, one tailed) is a function of this differential regression of MZ and DZ cotwin scores, and provides evidence for the heritable nature of reading disability.

The pattern of findings with respect to each of the individual tests is also consistent. The difference between the means for DZ probands and their cotwins exceeds that between MZ probands and their cotwins for each measure, and corresponding estimates of B_2 are significant for Reading Recognition ($P = 0.003$), Spelling ($P = 0.04$), and Digit Span ($P = 0.03$). It can also be seen from Table 1 that the average resemblance for members of MZ and DZ twin pairs (i.e., B_1) ranges from 0.50 to 0.77, each of which is significant.

Because the discriminant score and individual measures listed in Table 1 are correlated with general mental ability, an extended model that included full-scale IQ as a covariate was also fitted to the data. In general, results of this analysis are congruent with those from the fit of the basic regression model to the same data set. The B_2 estimates again are significant for the discriminant score, reading recognition, spelling and digit span, with associated P values of 0.003, 0.003, 0.04, and 0.05, respectively. Thus, the significance of the B_2 estimates presented in Table 1 is not attributable to the greater similarity in general mental ability of MZ compared with DZ twin pairs.

Although a genetic aetiology for reading disability is inferred from these results, it does not completely account for observed group differences between probands and controls. An estimate of the extent to which the group difference is heritable (h_g^2) could be obtained by dividing the estimate of B_2 by the difference between the proband mean and that of the unselected population. However, the unselected population mean is unknown because both the reading-disabled and control samples were selected on the basis of reading performance. Nevertheless, an approximate estimate of the unselected population mean was obtained from a weighted average of the control and proband means (95% and 5%, respectively) and then used to estimate h_g^2 . The resulting estimate for the discriminant score is 0.29 ± 0.10 , suggesting that about 30% of the reading deficit in probands arises from heritable factors. Thus, environmental differences contribute to the observed group differences. Furthermore, these results suggest that the performance deficits of reading-disabled children are not due solely to a fully penetrant major gene, though one or more of its forms could be inherited in a simple Mendelian manner⁴.

The flexibility of the regression analysis of selected twin data used here would permit testing for differential genetic aetiology by simultaneous analysis of data from probands of ostensibly different subtypes. Thus, if reading disability is a heterogeneous disorder¹³, such analysis could provide evidence for subtype validity. Moreover, by fitting an augmented model to the data, the extent to which individual differences in the unselected population are due to heritable influences (h^2) and to shared environmental influences (c^2) could be estimated⁵. h_g^2 and h^2 could differ in magnitude as a result of the aetiology of the deficit in probands differing from that of individual variations within the normal range. For example, deviant scores of probands could arise from environmental insult, whereas differences among individuals within the normal range may be multifactorial in origin. Thus, a comparison of h_g^2 and h^2 estimated from the same data set would provide a test of the hypothesis that probands merely represent the lower end of a normal distribution of individual differences¹⁴. These hypotheses will be tested in

the Colorado Reading Project when a larger sample of reading-disabled twins becomes available.

This research is supported in part by grants from the National Institute of Child Health and Human Development. We thank staff members of the Colorado school districts and the families who participated in the study.

Received 9 April; accepted 21 August 1987.

1. Brown, B. S. in *Dyslexia: An Appraisal of Current Knowledge* (eds Benton, A. L. & Pearl, D.) v-vii (Oxford University Press, 1978).
2. Critchley, M. in *The Dyslexic Child* (Heinemann, New York, 1970).
3. DeFries, J. C., Vogler, G. P. & LaBuda, M. C. in *Perspectives in Behavior Genetics* (eds Fuller, J. L. & Simmel, E. C.) 29-56 (Erlbaum, Hillsdale, 1986).
4. Smith, S. D. *Genetics and Learning Disabilities* (College-Hill, San Diego, 1986).
5. DeFries, J. C. & Fulker, D. W. *Behav. Genet.* **15**, 467-473 (1985).
6. LaBuda, M. C., DeFries, J. C. & Fulker, D. W. *Genet. Epidemiol.* **3**, 425-433 (1986).
7. DeFries, J. C. in *Biobehavioral Measures of Dyslexia* (eds Gray, D. B. & Kavanagh, J. F.) 107-122 (York, Parkton, 1985).
8. Wechsler, D. I. *Examiner's Manual: Wechsler Intelligence Scale for Children-Revised* (Psychological Corporation, New York, 1974).
9. Wechsler, D. I. *Examiner's Manual: Wechsler Adult Intelligence Scale-Revised* (Psychological Corporation, New York, 1981).
10. Dunn, L. M. & Markwardt, F. C. *Examiner's Manual: Peabody Individual Achievement Test* (American Guidance Service, Circle Pines, 1970).
11. DeFries, J. C., Plomin, R., Vandenberg, S. G. & Kuse, A. R. *Intelligence* **5**, 245-277 (1981).
12. Nichols, R. C. & Bilbro, W. C. *Acta Genetica* **16**, 265-275 (1966).
13. Satz, P., Morris, R. & Fletcher, J. M. in *Biobehavioral Measures of Dyslexia* (eds Gray, D. B. & Kavanagh, J. F.) 25-40 (York, Parkton, 1985).
14. Rodgers, B. *Br. J. educ. Psychol.* **53**, 369-373 (1983).

Two forms of transforming growth factor- β distinguished by multipotential haematopoietic progenitor cells

Masatsugu Ohta*, Joel S. Greenberger*,
Pervin Anklesaria*, Anna Bassols† & Joan Massagué†

Departments of * Radiation Oncology and † Biochemistry,
University of Massachusetts Medical School, Worcester,
Massachusetts 01605, USA

Type- β transforming growth factors (TGF- β s) are polypeptides that act hormonally to control proliferation and differentiation of many cell types^{1,2}. Two distinct homodimeric TGF- β polypeptides, TGF- β 1 and TGF- β 2 have been identified which show ~70% amino-acid sequence similarity^{3,4}. Despite their structural differences, TGF- β 1 and TGF- β 2 are equally potent at inhibiting epithelial cell proliferation and adipogenic differentiation³. The recent immunohistochemical localization of high levels of TGF- β in the bone marrow and haematopoietic progenitors of the fetal liver⁵ has raised the possibility that TGF- β s might be involved in the regulation of haematopoiesis. Here we show that TGF- β 1, but not TGF- β 2, is a potent inhibitor of haematopoietic progenitor cell proliferation. TGF- β 1 inhibited colony formation by murine factor-dependent haematopoietic progenitor cells in response to interleukin-3 (IL-3) or granulocyte-macrophage colony stimulating factor (GM-CSF), as well as colony formation by marrow progenitor cells responding to CSF-1 (M-CSF). The progenitor cell lines examined were ~100-fold more sensitive to TGF- β 1 than TGF- β 2, and displayed type-I TGF- β receptors with affinity ~20-fold higher for TGF- β 1 than TGF- β 2. These results identify TGF- β 1 as a novel regulator of haematopoiesis that acts through type-I TGF- β receptors to modulate proliferation of progenitor cells in response to haematopoietic growth factors.

Murine IL-3 (multi-CSF) supports the proliferation of multipotential progenitor cells as well as granulocyte, macrophage, erythroid, megakaryocyte and mast cells⁶. The effect of TGF- β s on haematopoietic progenitor cell proliferation was studied using both fresh haematopoietic cells and IL-3-dependent haematopoietic cell lines^{7,8}. The B6SutA cell line forms multilineage granulocyte-macrophage erythroid-mast cell colonies as well as macroscopic haemoglobinized erythroid bursts in response to IL-3 and erythropoietin in semisolid medium (Fig.

1). Picomolar TGF- β 1 markedly (~70%) inhibited multilineage as well as haemoglobinized colony formation by B6SutA cells (Fig. 1). A significant but less extensive inhibition of colony formation by TGF- β 1 was observed in 32D-C13 cells, an IL-3-dependent bilineage progenitor cell line which differentiates to mast cells in response to IL-3, or granulocytes in response to G-CSF (refs 7 and 9). In contrast, TGF- β 2 at concentrations up to 1 nM did not affect colony formation by either cell line (Fig. 1) even though it acts at picomolar concentrations on other mouse cell types³.

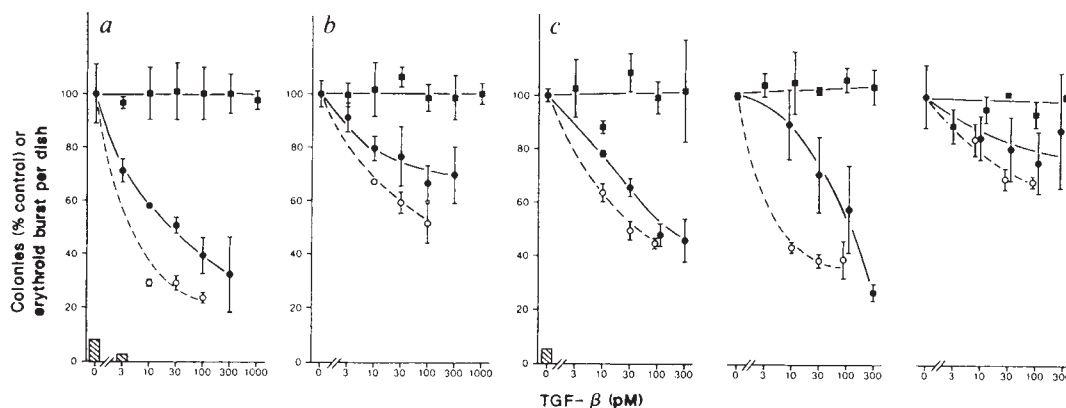
The formation of colonies by freshly isolated bone marrow cells in response to IL-3, GM-CSF (which stimulates production of granulocytes and macrophages)¹⁰, and CSF-1 (which preferentially supports the proliferation and development of cells of the monocyte/macrophage lineage)¹¹ was inhibited $37 \pm 1\%$, $21 \pm 2\%$ and $10 \pm 2\%$ by 0.1 nM TGF- β 1, respectively (not shown). To reduce the possibility that marrow stromal cells or other accessory cells interfered with the effects of TGF- β 1, the experiments were repeated with non-adherent haematopoietic progenitors from murine long-term bone marrow cultures (LTBMCs)¹². TGF- β 1 inhibited strongly the formation of colonies induced by IL-3 or GM-CSF and to a lesser extent by L-cell conditioned medium¹¹ used as a source of CSF-1 (Fig. 1). Again, TGF- β 2 was ineffective in regulating colony formation by haematopoietic progenitors from LTBMCs (Fig. 1). TGF- β radioreceptor assays¹³ of culture supernatants showed no significant difference between the rates of degradation of TGF- β 1 and TGF- β 2 which were 20-25% per day when the factors were added at 0.5 nM to suspension cultures (10^6 cells ml⁻¹) of B6SutA cells. TGF- β 1 added three days after plating was still capable of inhibiting colony formation (Fig. 1), but cells eventually became refractory to TGF- β 1 as no inhibition of further colony formation was observed when TGF- β 1 was added on day 7 (data not shown).

The effect of TGF- β on the proliferative response of B6SutA and 32D-C13 cells to IL-3 was also examined in suspension cultures. TGF- β 1 induced a marked decrease in thymidine incorporation into macromolecules in both cell lines (Fig. 2). In contrast, proliferation of B6Sut-C127 cells, an IL-3-independent B6SutA-derived clone, was not affected by TGF- β 1 (Fig. 2) even though these cells have apparently normal TGF- β receptors. As in the inhibition of haematopoietic colony formation, TGF- β 2 was much less potent than TGF- β 1 in exerting an antiproliferative response on B6SutA cells and 32D-C13 cells in suspension culture. In this assay, TGF- β 2 exhibited at most ~1% of the potency of TGF- β 1 (Fig. 2). These results suggest that the TGF- β s control proliferation of the haematopoietic cells examined by interaction with receptors that can discriminate between TGF- β 1 and TGF- β 2.

Two of the three types of cell surface receptors for TGF- β identified in many cells by affinity labelling are glycoproteins of relative molecular mass (M_r) ~53,000 (53 K) (type-I receptors) and ~73 K (type-II receptors), and both have higher affinity for TGF- β 1 than for TGF- β 2 (refs 3 and 14). The type-III TGF- β receptor is a large (~300 K) glycoprotein¹⁵ which has an affinity for TGF- β 1 similar to that for TGF- β 2 (ref. 3). Occupancy of type-III receptors with TGF- β 1 or TGF- β 2 has been implicated in the mediation of certain responses to TGF- β s, including inhibition of epithelial cell proliferation, inhibition of adipogenesis, and induction of cell adhesion protein expression, because TGF- β 1 and TGF- β 2 are equally potent at eliciting these responses³. Despite the high affinity of type I and type II receptors for TGF- β 1, no biological actions had previously been ascribed to them.

Analysis¹⁶ of ¹²⁵I-labelled TGF- β 1 binding to the haematopoietic cell lines indicated the presence of a single class of high-affinity binding sites (240 sites per cell with dissociation constant $K_d = 9$ pM in B6SutA; 110 sites per cell with $K_d = 19$ pM in B6Sut-C127; 270 sites per cell with $K_d = 16$ pM in 32D-C13) (not shown). Affinity labelling of B6SutA cells by

Fig. 1 Response of haematopoietic progenitor cells (*a*, B6SutA; *b*, 32DC13; *c*, LTMBc progenitors) to TGF- β 1 and TGF- β 2 in the presence of colony stimulating factors as follows. *a*, IL-3 and erythropoietin; *b*, IL-3 alone; *c*, IL-3 and erythropoietin (left), GM-CSF (centre) and L-cell conditioned medium (right). Progenitor cell assays²¹ were done with murine factor-dependent haematopoietic progenitor cell lines B6SutA and 32D-C13, and non-adherent cells from murine long-term bone marrow cultures (LTBMCs)¹² in the presence of colony stimulating factors. Various concentrations of human platelet TGF- β 1 (refs 3 and 22) (●, and hatched bars) or porcine platelet TGF- β 2 (ref. 3) (R&D Systems) (■) were added at the beginning of the assay. In some experiments TGF- β 1 (○), dissolved in 0.1 ml of medium, was added on the surface of the methylcellulose cultures 3 days after plating. Cultures were incubated at 37 °C and colonies of >50 cells (symbols) as well as erythroid cell bursts (hatched bars) were scored at day 7 or day 14. Results are mean $\pm$ s.d. of duplicate plates from three experiments and are expressed as the percentage of the values obtained in cultures without TGF- β 1.



Methods. B6SutA and 32D-C13 cells recloned from single cells were then maintained in McCoy's 5A medium supplemented with 15% fetal bovine serum (Gibco) and 10% WEHI-3B conditioned medium⁶ as a source of IL-3. LTBMCs were established from B6D2F1 mice. Non-adherent cells harvested at day 24 or 40 were used as a source of non-adherent haematopoietic progenitor cells. In *a*, B6SutA cells were plated with 2.5 ng ml⁻¹ of murine recombinant IL-3 (rIL-3, Biogen) and 2.5 U ml⁻¹ erythropoietin (Epo, Terry Fox Laboratories, Vancouver), and formed 91.4 $\pm$ 10.2 multilineage granulocyte-erythroid-macrophage-megakaryocyte (CFU-GEMM) colonies per 5 $\times$ 10³ cells, in the absence of TGF- β 1; in *b* 32D-C13 cells were plated with 2.5 ng ml⁻¹ IL-3 and formed 155.3 $\pm$ 7.6 mast cell colonies per 1 $\times$ 10³ cells in the absence of TGF- β 1; and in *c*, purified haematopoietic progenitor cells plated with 2.5 ng ml⁻¹ rIL-3 and 2.5 U ml erythropoietin formed 59.0 $\pm$ 1.4 CFU-GEMM colonies per 5 $\times$ 10⁴ cells; with 1.0 ng ml⁻¹ murine recombinant GM-CSF (rGM-CSF, Biogen) they formed 48.5 $\pm$ 0.7 granulocyte-macrophage (CFU-GM) colonies per 5 $\times$ 10⁴ cells and with 10% L-cell conditioned medium as a source of CSF-1 they formed 55.0 $\pm$ 6.5 macrophage (CFU-M) colonies per 5 $\times$ 10⁴ cells, in the absence of added TGF- β .

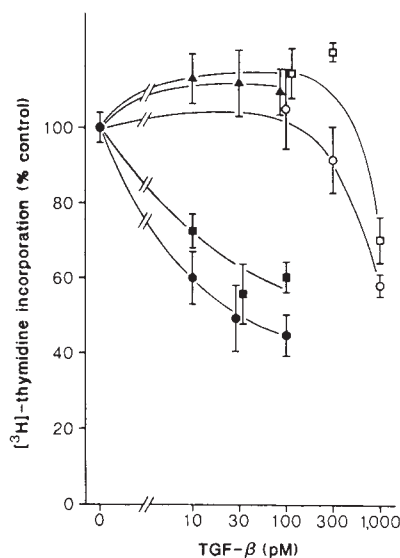


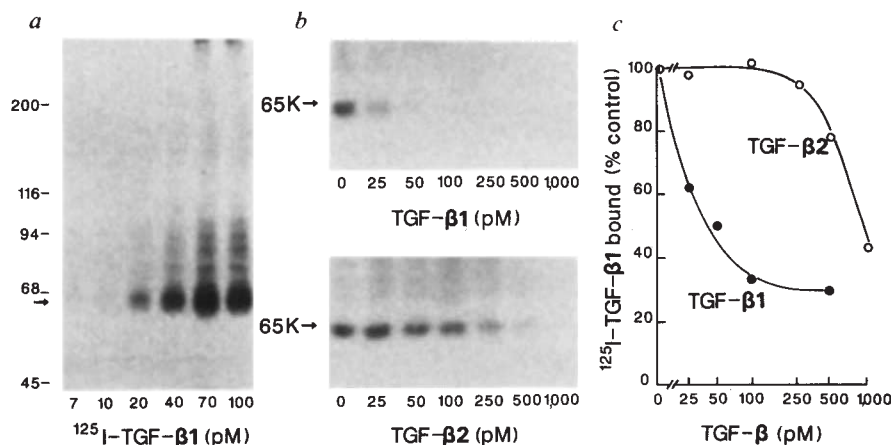
Fig. 2 Effect of TGF- β 1 or TGF- β 2 on proliferation of factor-dependent haematopoietic cell lines. Samples of B6SutA cells (●, ○) 32D-C13 cells (■, □) or B6Sut-C127 (▲) (10⁴ cells ml⁻¹, with 2.5 ng ml⁻¹ rIL-3 in the B6SutA and 32D-C13 cultures) were incubated for 48 h at 37 °C with the indicated concentration of TGF- β 1 (closed symbols) or TGF- β 2 (open symbols). The cells were pulsed for 24 h with 1 μ Ci per well of ³H-thymidine, harvested and precipitated with 10% trichloroacetic acid, and the precipitate collected on a filter paper. The incorporated radioactivity was determined by liquid scintillation counting. The results are mean $\pm$ s.d. of triplicate determination from three experiments, and are expressed as per cent of the values obtained in cultures without TGF- β . Radioactivity incorporated by B6SutA and 32D-C13 cells in the absence of IL-3 and TGF- β was 0.1% and 1.4% of the values (82.5 $\pm$ 6.2 $\times$ 10³ c.p.m. and 99.6 $\pm$ 5.4 $\times$ 10³ c.p.m., respectively) in the presence of rIL-3.

cross-linking to receptor-bound ¹²⁵I-labelled TGF- β 1, followed by electrophoresis and autoradiography of the gels, revealed a single predominant labelled band of 65 K (Fig. 3*a* and *b*). This labelled receptor species (of $\sim$ 53 K after subtracting the 12 K contributed by the crosslinked TGF- β monomer) corresponds in size to type-I TGF- β receptors observed in many other cell types. The labelling of this TGF- β receptor species was half maximal at $\sim$ 20 pM ¹²⁵I-labelled TGF- β 1, and maximal at $\sim$ 70 pM ¹²⁵I-labelled TGF- β 1 (Fig. 3*a*), in good agreement with the ligand binding data. Unlabelled TGF- β 1 at 25 pM inhibited half-maximally the labelling of the receptor by 25 pM ¹²⁵I-labelled TGF- β 1 (Fig. 3*b*). TGF- β 2 was 15–20 times less potent than TGF- β 1 at competing with ¹²⁵I-labelled TGF- β 1 for binding to B6SutA cells (Fig. 3*c*) and at inhibiting affinity labelling of the receptor (Fig. 3*b*). Affinity labelling of B6Sut-C127 and 32D-C13 cells with ¹²⁵I-labelled TGF- β 1 and competition experiments with TGF- β 1 and TGF- β 2 showed that the TGF- β receptors in these two cell lines had the same molecular mass and ligand recognition properties as the B6SutA line (not shown). The ligand recognition properties of the 65 K labelled receptor are characteristic of type-I TGF- β receptors, and allow us to conclude that occupancy of this receptor type in B6SutA and 32D-C13 cells correlates with inhibition of haematopoietic progenitor cell proliferation.

These studies show that TGF- β 1 is a strong inhibitor of haematopoietic progenitor cell proliferation in response to colony stimulating factors. Thus, TGF- β 1 may be critical in balancing the development of the haematopoietic system *in vivo*. As postulated in general terms before¹⁷, a mechanism for transformation of haematopoietic progenitor cells could be the escape from negative regulation by TGF- β , which could act together with unregulated CSF production^{18–20}.

The therapeutic potential of TGF- β has been discussed¹. However, the wide range of cell types whose proliferation and differentiation can be potentially disturbed by TGF- β raises questions about the mechanisms that may restrict TGF- β action. The existence of different forms of this factor (TGF- β 1, TGF- β 2

Fig. 3 TGF- β receptors in B6SUtA cells. *a*, Autoradiogram from a SDS polyacrylamide electrophoresis gel containing samples from B6SUtA cells affinity labelled with various concentrations of human 125 I-labelled TGF- β 1. Affinity labelling was performed by crosslinking 2×10^6 cells in 1 ml of binding buffer with the receptor-bound radioligand using disuccinimidyl suberate as previously described²³. Arrow, prominent 65 K affinity-labelled species; markers, M_r in thousand. *b*, Autoradiograms from gels containing samples from cells affinity labelled with 25 pM 125 I-labelled TGF- β 1 alone or with the indicated concentrations of TGF- β 1 or TGF- β 2. *c*, Binding competition curves obtained with B6SUtA cells incubated with 25 pM 125 I-labelled TGF- β 1 alone or with increasing concentrations of TGF- β 1 or TGF- β 2. Binding assays using cell suspensions were done at 0–4 °C as previously described^{13,23}.



and the heterodimer, TGF- β 1.2)³ in conjunction with the discriminating ligand binding properties of TGF- β receptor types I and II may provide the opportunity to influence selectively certain cellular functions but not others. *In vivo* TGF- β 2, from endogenous sources or exogenously administered, may mimic potentially only those actions of TGF- β 1 that are exerted through receptors shared by both factors, whereas it may act as a poor agonist on systems which, like haematopoietic progenitors, are regulated by receptors that bind TGF- β 1 preferentially.

We thank S. Cheifetz for 125 I-labelled TGF- β and valuable discussions. This work was supported by NIH grants to J.S.G. and J.M. A.B. is the recipient of a postdoctoral fellowship from Spanish Ministerio de Educacion y Ciencia/Fulbright Foundation.

Received 14 May; accepted 14 August 1987.

1. Sporn, M. B., Roberts, A. B., Wakefield, L. M. & Assoian, R. K. *Science* **233**, 532–534 (1986).
2. Massagué, J. *Cell* **49**, 437–438 (1987).

3. Cheifetz, S. *et al. Cell* **48**, 409–415 (1987).
4. Seyedin, S. M. *et al. J. biol. Chem.* **262**, 1946–1949 (1987).
5. Ellingsworth, L. R. *et al. J. biol. Chem.* **261**, 12362–12367 (1986).
6. Ihle, J. N. *et al. J. Immun.* **131**, 282–287 (1983).
7. Greenberger, J. S., Sakakeeny, M. A., Humphries, R. K., Eaves, C. J. & Eckner, R. J. *Proc. natn. Acad. Sci. U.S.A.* **80**, 2931–2935 (1983).
8. Greenberger, J. S. *et al. Fedn Proc.* **42**, 2762–2771 (1983).
9. Valtieri, M. *et al. J. Immun.* (in the press).
10. Metcalf, D. *Blood* **67**, 257–267 (1986).
11. Stanley, E. R. & Heard, P. M. *J. biol. Chem.* **252**, 4305–4312 (1977).
12. Greenberger, J. S. *Nature* **275**, 752–754 (1978).
13. Massagué, J. *Meth. Enzym.* **146**, 174–195 (1987).
14. Cheifetz, S., Like, B. & Massagué, J. *J. biol. Chem.* **261**, 9972–9978 (1986).
15. Massagué, J. *J. biol. Chem.* **260**, 7059–7066 (1985).
16. Scatchard, G. *Ann. N.Y. Acad. Sci.* **51**, 660–672 (1949).
17. Sporn, M. B. & Roberts, A. B. *Nature* **313**, 745–747 (1985).
18. Schrader, J. W. & Crapper, R. M. *Proc. natn. Acad. Sci. U.S.A.* **80**, 6892–6896 (1983).
19. Lang, R. A., Metcalf, D., Gough, N. M., Dunn, A. R. & Gonda, T. J. *Cell* **43**, 531–542 (1985).
20. Hapel, A. J., Lee, J. C., Farrar, W. L. & Ihle, J. N. *Cell* **25**, 179–186 (1981).
21. Metcalf, D. *The Haemopoietic Colony Stimulating Factors* 229–275 (Elsevier, Amsterdam, 1984).
22. Assoian, R. K., Komoriya, A., Meyers, C. A., Miller, D. M. & Sporn, M. B. *J. biol. Chem.* **258**, 7155–7160 (1983).
23. Massagué, J. & Like, B. *J. biol. Chem.* **260**, 2636–2645 (1985).

T-cell receptors of human suppressor cells

Robert L. Modlin*, Michael B. Brenner†‡, Michael S. Krangel†, Allan D. Duby§ & Barry R. Bloom||

* Section of Dermatology and Department of Pathology, University of Southern California School of Medicine, Los Angeles, California 90033, USA

† Division of Tumor Virology, Dana-Farber Cancer Institute, Harvard Medical School, Boston, Massachusetts 02115, USA

‡ Department of Rheumatology and Immunology, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts 02115, USA

§ Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA

|| Department of Microbiology and Immunology, Albert Einstein College of Medicine, Bronx, New York, 10461, USA

Cells which can suppress the immune response to an antigen (T_s cells) appear to be essential for regulation of the immune system^{1,2}. But the characterization of the T_s lineage has not been extensive and many are sceptical of studies using uncloned or hybrid T-cell lines. The nature of the antigen receptor on these cells is unclear. T cells of the helper or cytotoxic lineages appear to recognize their targets using the T-cell receptor (TCR) $\alpha\beta$ -CD3 complex^{3–5}. TCR β -gene rearrangements are also found in some murine and human suppressor cell lines^{6–10} but others have been shown not to

rearrange or express the β -chain^{11–13} or α -chain¹⁴ genes. We previously established T_s clones derived from lepromatous leprosy patients^{15,16} which carry the CD8 antigen and recognize antigen in the context of the major histocompatibility complex (MHC) class II molecules *in vitro*¹⁶. We here report the characterization of additional MHC-restricted T_s clones which rearrange TCR β genes, express messenger RNA for the α and β chains of the TCR and express clonally unique CD3-associated TCR $\alpha\beta$ structures on their cell surface but do not express the γ chain of the $\gamma\delta$ TCR on the cell surface. We conclude that antigen recognition by at least some human CD8⁺ suppressor cells is likely to be mediated by TCR $\alpha\beta$ heterodimers.

Leprosy provides an extraordinary model for investigating immunoregulatory mechanisms in man^{17,18}. The disease exhibits a clinical spectrum in which the stage of disease can be correlated with the cellular immune responses of the patients to antigens of *Mycobacterium leprae*. In lepromatous leprosy, patients exhibit a selective and specific unresponsiveness to *M. leprae* antigens. We have suggested that this unresponsiveness may partly result from T cells that recognize *M. leprae*-specific antigens or epitopes and block development of potentially beneficial responses of T_H cells *in vivo*¹⁹. Such *M. leprae*-specific T_s cells have been found in the blood of lepromatous patients and characterized as CD3⁺CD8⁺DR⁺ and FcR⁺ (refs 20, 21). Using CD8 cells cloned directly from skin lesions of lepromatous patients and the recently developed antibody reagents against framework determinants on human TCR α , β and γ polypeptides^{22–26}, it became possible directly to examine the antigen receptors of human T_s cells in molecular terms.

T_s clones were derived from skin biopsies of lepromatous

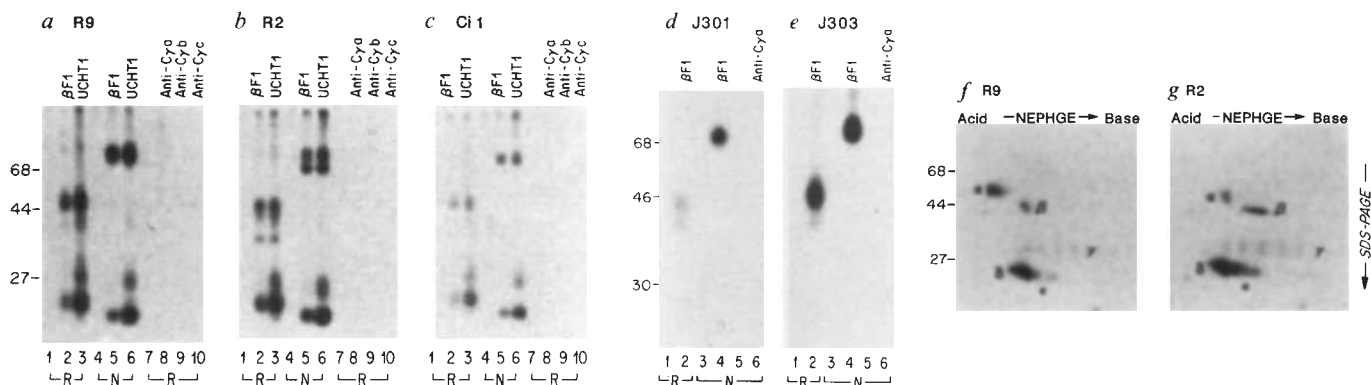


Fig. 1 Immunoprecipitation of T-cell receptors from T_S clones. 125 I-labelled R9 (a), R2 (b), Ci1 (c), J301 (d) and J303 (e) lymphocytes were detergent solubilized and immunoprecipitated with anti-CD3 or anti-TCR antibodies and analysed by one- or two-dimensional gel analyses. a, b and c, One-dimensional SDS-PAGE analysis. Immunoprecipitations were performed on CHAPS solubilized R9, R2 or Ci1 T_S lymphocytes using mAb P3 (lanes 1 and 4), mAb β F1 (lanes 2 and 5), anti-CD3 mAb UCHT1 (lanes 3 and 6). Lanes 1–3 were analysed under reducing conditions (R), while lanes 4–6 were aliquots of the same immunoprecipitates analysed under nonreducing conditions (N). Immunoprecipitations from denatured (1% SDS) cell lysates were performed with control normal rabbit serum (NRS) (lane 7), rabbit anti- $C_{\gamma a}$ sera (lane 8), rabbit anti- $V_{\gamma b}$ (lane 9) and rabbit anti- $C_{\gamma c}$ (lane 10) all carried out under reducing conditions. d, e, Immunoprecipitates were performed on TX-100 solubilized J301 or J303 T_S lymphocytes using mAb P3 (lanes 1 and 3) or mAb β F1 (lanes 2 and 4) and analyses were carried out under R or N conditions. Immunoprecipitates from denatured (1% SDS) cell lysates were performed with control normal rabbit serum (lane 5) and with rabbit anti- $C_{\gamma a}$ (lane 6) under N conditions. f, g, Two-dimensional gel analyses of T_S T-cell receptors. Immunoprecipitates from 125 I-labelled CHAPS solubilized R9 (f) or R2 (g) lymphocytes were performed using anti-CD3 mAb UCHT1. The precipitates were separated by charge in the first dimension (NEPHGE) under R conditions followed by size separation (SDS-PAGE) in the second dimension, under R conditions.

Methods. One-dimensional SDS-PAGE. Viable lymphocytes were isolated by Ficoll-Hypaque density gradient centrifugation and 1×10^7 cells were radioiodinated by the lactoperoxidase technique as previously described⁴³. Labelled cells were solubilized in 5 ml of TBS (10 mM Tris pH 8, 140 mM NaCl) containing 2 mM phenylmethylsulphonyl fluoride (PMSF) and 8 mM iodoacetamide (IAA), with 0.3% 3-[(3-cholamidopropyl) dimethylammonio]-1-propane-sulphonate (CHAPS) (a, b, c, f and g which preserve the TCR-CD3 association)⁴⁴ or with 1% Triton X-100 which dissociates the CD3-TCR complex (d, e). Immunoprecipitation was carried out using fixed *Staphylococcus aureus* Cowan I (SACI) as previously described³², and the immune complexes were washed five times in TBS containing 0.1% Triton X-100 (TX-100). Rat anti-mouse κ chain-specific mAb 187.1 (5 μ g) was added as a second antibody to provide protein A binding of IgG, mAb β F1, UCHT1 and P3⁴⁵. Reduced samples were boiled in 2 mM dithiothreitol (DTT) and all samples incubated for 10 min at 23 °C in 10 mM IAA before analysis by SDS-PAGE. Immunoprecipitations using anti- C_{γ} sera were performed on 125 I-labelled T_S cells that were solubilized in 1% TX-100, dialysed to remove IAA, and then denatured in 1% SDS containing 3 mM DTT by boiling for 3 min. After partial renaturation by the addition of 4 volumes of 1.5% TX-100 in TBS containing 30 mM IAA, anti- C_{γ} sera or NRS were added, followed by SACI, and the immunoprecipitates were washed in TBS containing 0.5% TX-100, 0.5% deoxycholate, 0.05% SDS before analysis by SDS-PAGE. Anti- $C_{\gamma a}$ sera has been previously reported²⁴, anti- $V_{\gamma b}$ was generated against a 17 amino-acid synthetic peptide (Arg-Thr-Lys-Ser-Val-Thr-Arg-Gln-Thr-Gly-Ser-Ser-Ala-Glu-Ile-Thr-Cys) and anti- $C_{\gamma c}$ sera was generated against a 12 amino-acid synthetic peptide (Val-Pro-Glu-Lys-Ser-Leu-Asp-Lys-Glu-His-Arg-Cys), a gift from T Cell Sciences, Inc. (Cambridge, MA). For two-dimensional gel analysis, 125 I-labelled T_S cells were immunoprecipitated with UCHT1 as in a. NEPHGE (charge separation) was then carried out using pH 3.5–10 ampholines (LKB, Sweden), followed by 10.5% SDS-PAGE gels for size separation as previously described²⁶. All gels were exposed to Kodak XAR film for autoradiography.

leprosy lesions following cytofluorographic cell sorting for CD8⁺ cells as described previously^{15,16,21}. To avoid the possibility that mitogens might bind to the CD3-TCR complex and selectively expand cells with a certain type of receptor or otherwise bias the results, all clones were initially established by the method of limiting dilution cloning with propagation in the presence of interleukin-2 (IL-2) alone. The phenotypes and antigen-induced suppressor activity of these clones are summarized in Table 1. All the leprosy-derived lymphocyte clones were CD3⁺8⁺4⁺ and WT31⁺. These clones exhibited lepromin-induced suppression of the concanavalin A (Con A) response of peripheral blood mononuclear cells from normal donors as previously described¹⁶. Moreover, in five of five clones tested, the suppressor activity observed appeared to be MHC-restricted (Table 1 and ref. 16). On the basis of preliminary cytotoxicity data using lepromin-treated, MHC-compatible target cells that included human B cell lines, ori-sarcoma virus 40 transformed macrophage lines and 6-day primary cultured monocytes, we could find no evidence that the mechanism of suppressor activity of these clones involved cytotoxicity of antigen presenting cells (data not shown). Similar findings were obtained by De Vries *et al.*²⁷.

In light of the known linkage between the cell-surface expression of CD3 and TCR $\alpha\beta$ ^{28–32}, or CD3 and TCR $\gamma\delta$ ²⁴ subunits, we undertook an investigation of CD3-associated TCR

proteins on our T_S clones. T_S clones R9, R2 and Ci1 were surface iodinated, detergent solubilized (under conditions that preserve the CD3-TCR association), and labelled membrane proteins were immunoprecipitated using the anti-CD3 monoclonal antibody, UCHT1. In addition to the CD3 subunits (M_r 20,000–27,000 (20–27 K)), CD3-associated polypeptides were detected on all T_S clones (Fig. 1). When examined under reducing conditions, R9 and R2 cells expressed two CD3-associated species of 40–50K, while only one species was detected on Ci1 cells (Fig. 1a, b, c, lane 3 for each clone). When aliquots of these same immunoprecipitates were compared under nonreducing conditions, all of the CD3-associated species were noted to be disulphide-linked (Fig. 1a, b, c, lane 6 for each clone). Thus, human T_S clones R9, R2 and Ci1 express 40–50 K disulphide-linked, CD3-associated polypeptides.

To determine whether these species represented the protein products of TCR α or β genes, reactivity with framework monoclonal antibodies specific for the TCR $\alpha\beta$ complex was examined. The WT31 monoclonal antibody preferentially reacts with TCR $\alpha\beta$ ⁺ lymphocytes in cytofluorographic analysis²³. All of the T_S clones exhibited positive staining with this mAb, suggesting that TCR $\alpha\beta$ proteins were expressed on their cell surfaces (Table 1). However, this monoclonal antibody is inefficient in immunoprecipitation studies²³ such that it is often not possible to isolate surface proteins directly using this reagent.

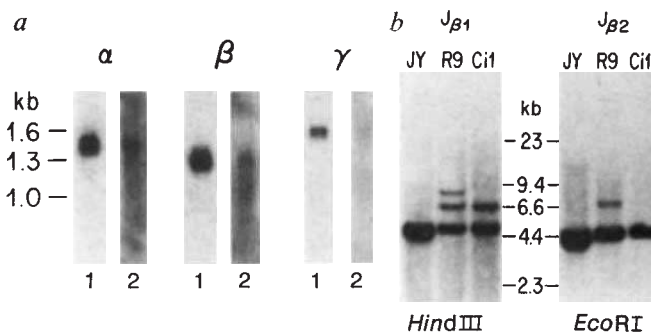


Fig. 2 Northern and Southern blot analysis of R9 RNA and DNA. **a**, Northern analysis. Total RNA preparations from the T leukaemic cell line HPB-MLT (lane 1 for each probe) or from T_S clone R9 (lane 2 for each probe) were analysed on Northern blots using TCR α , TCR β and TCR γ cDNA probes. **b**, Southern analysis of TCR β gene rearrangements in R9 and Ci1 cells. Genomic DNA was isolated from a B lymphoblastoid cell line, JY (germline), and from T_S cells, R9 and Ci1. Genomic DNAs were digested with *Hind*III (Fig. 2b, $J_{\beta 1}$ probe) or with *Eco*RI (Fig. 2b, $J_{\beta 2}$ probe) and hybridized with ³²P-labelled $J_{\beta 1}$ and $J_{\beta 2}$ probes as indicated. **Methods.** **a**, RNA preparation, agarose gel electrophoresis, transfer to nitrocellulose, hybridization with ³²P-labelled, nick translated probes, and autoradiography were as described previously²⁴. Approximately 0.5–1.0 μ g of RNA was loaded per lane, probes were labelled to similar specific activity, and identical times of autoradiographic exposures are presented. RNA sizes were determined based on previously published lengths for TCR β and TCR γ transcripts^{35,46}. **b**, Genomic DNA was isolated as previously described⁴⁷, digested with restriction enzymes *Hind*III or *Eco*RI, size fractionated on 0.7% agarose gels and transferred to nitrocellulose filters for hybridization with ³²P-labelled $J_{\beta 1}$ or $J_{\beta 2}$ probes, respectively. The organization of the human TCR β gene has been reported⁴⁸. Probes are restriction enzyme DNA fragments of previously isolated human genomic clones⁴⁹. $J_{\beta 1}$ is a 2.6-kilobase (kb) *Hind*III–*Nsi*I fragment containing the $J_{\beta 1}$ gene segment cluster. Because of a *Hind*III site in the 5' end of $J_{\beta 1.1}$, the most 5' of the $J_{\beta 1}$ gene segments, rearrangements to this particular segment may not be detected on Southern blot analyses of *Hind*III-digested genomic DNA. $J_{\beta 2}$ is a 4.4-kb *Eco*RI–*Eco*RI fragment containing the $J_{\beta 2}$ gene segment cluster. This probe detects rearrangements to the $J_{\beta 2}$ gene segment cluster on Southern blot analyses of *Eco*RI-digested DNA. The $J_{\beta 1}$ and $J_{\beta 2}$ probes do not cross-hybridize. After hybridization, filters were washed in 2 $\times$ SSC and 0.1% SDS at 23 °C followed by 0.2 $\times$ SSC and 0.1% SDS at 68 °C before autoradiography with intensifying screens. Because of the varying contribution of DNA from irradiated B lymphoblastoid cells used as a feeder layer, the 4.0 kb *Hind*III fragment containing the germline $J_{\beta 1}$ gene segment and the 4.4-kb *Eco*RI fragment containing the germline $J_{\beta 2}$ gene segments are present even in lanes containing two rearranged bands. The two filters were first hybridized to a TCR β constant region (C_{β}) cDNA probe which failed unequivocally to reveal rearrangements. The filters were therefore stripped in 0.5 M NaOH before the hybridizations shown (above) with the genomic $J_{\beta 1}$ and $J_{\beta 2}$ probes.

Thus, an additional monoclonal antibody (mAb), β F1, which has framework reactivity against the TCR β subunit and is capable of immunoprecipitating the disulphide-linked TCR $\alpha\beta$ complex from the surface of T lymphocytes²⁵ was used. This antibody immunoprecipitated polypeptides (40–50 K reduced; 80–90 K nonreduced) which appeared identical to those co-immunoprecipitated using anti-CD3 mAb (Fig. 1a, b, c lanes 2, 5 for each clone). Similar results were obtained using mAb β F1 for immunoprecipitation from radiolabelled Triton-X 100 solubilized J301 and J303 T_S cells. Under these conditions, CD3 and TCR do not remain associated such that the TCR α and β are isolated without CD3 (Fig. 1d, e). A second CD3-associated subunit could not be visualized by one-dimensional SDS-polyacrylamide gel electrophoresis (PAGE) on Ci1 or J303 cells either because this species was poorly radiolabelled with ¹²⁵I or because both species were of similar SDS-PAGE mobility; and

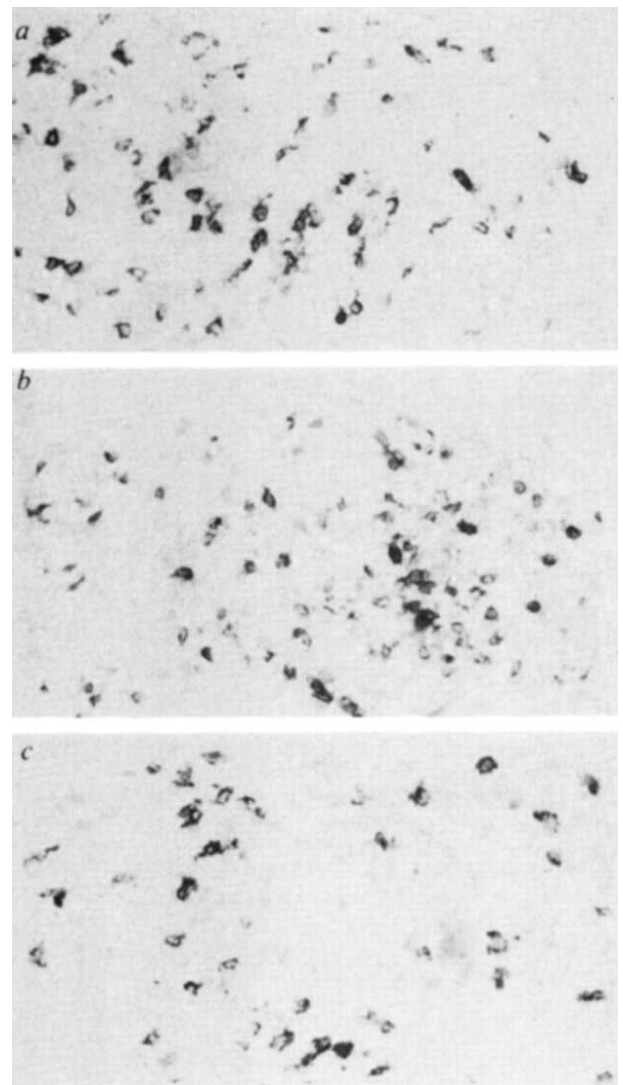


Fig. 3 Cells expressing CD3 and TCR β chains in lesions of lepromatous leprosy. In the serial frozen sections of a skin biopsy specimen from lepromatous leprosy patient R (from whom 2 of the 3 clones were derived, namely R9 and R2), approximately 40% of the infiltrating cells stained positively for CD3 (**a**, dark black rings are positive cells), 40% for TCR β chains (**b**) and 25% for CD8 (**c**) ($\times 130$). Double staining of CD8⁺ cells in this lesion showed that >80% were CD3⁺ and TCR β ⁺ (data not shown).

Methods. Frozen sections of skin biopsy specimens from 14 lepromatous leprosy patients were stained by indirect immunoperoxidase and counterstained with haematoxylin⁴⁰ with a panel of monoclonal antibodies CD3 (Leu4, Becton Dickinson), CD8 (Leu2a) and TCR β chains (β F1) as described. Double staining of CD8⁺ cells with CD3 and β F1 was performed on five lepromatous leprosy biopsy specimens using a double labelling technique. The first antibody was visualized with immunoperoxidase and aminoethylcarbazole to give a red reaction product and the second antibody was visualized with glucose oxidase showing a blue reaction product. The percentage of double staining cells could then be counted over several oil immersion fields.

two species were visualized under nonreducing conditions from R2 cells. When the CD3-associated species expressed on T_S clones R9 and R2 were examined by two-dimensional gel electrophoresis (nonequilibrium pH gradient electrophoresis (NEPHGE) followed by SDS-PAGE) under reducing conditions, clone-specific acidic 45–50 K species and basic 40–45 K species were observed (Fig. 1f, g). Thus, R2 and R9 displayed clonally unique, β F1 reactive, heterodimeric CD3-associated subunits characteristic of TCR $\alpha\beta$ complexes.

Table 1 Characterization of T_S clones

Clone	Phenotype (per cent)				Suppression (per cent)	
	CD3	CD4	CD8	WT31	MHC compat.	incompat.
R2	99	0	99	99	28	NT
R9	99	2	96	99	50	NT
Ci1	99	3	98	99	55	NT
J31	91	0	99	89	66	-8
J301	100	1	96	99	30	-13
J303	93	0	97	93	44	-32
J107	94	7	89	98	25	-10
J109	94	3	99	100	31	-3

Skin biopsy specimens were obtained from three lepromatous leprosy patients, R, Ci and J, and CD8⁺ clones were established directly from lesion-derived cells by limiting dilution in the presence of irradiated feeder cells and IL-2 as previously described^{15,16}. Clones R2 and R9 were maintained with irradiated feeder cells or alternately with IL-2 alone (Electronucleonics Inc.) or IL-2 plus phytohaemagglutinin (PHA, Leucoagglutinin, Pharmacia) plus an autologous EBV transformed B-cell line⁴¹. Clone Ci1 and all J clones were maintained with irradiated feeder cells; or alternatively with IL-2 alone; or IL-2 plus conditioned medium containing T-suppressor cell growth factor⁴². Phenotypes were determined by staining clones with FITC-conjugated CD3, CD4 and CD8 (Coulter Inc.) or WT31 followed by FITC conjugated goat anti-mouse IgG and analysing with a FACS IV (Becton Dickinson). Suppressor activity was determined by measuring lepromin-induced suppression of the Con A response of normal peripheral blood mononuclear cells. Responding cells were MHC matched (compat.) for at least one class I and one class II antigen for assays of J clones, or were matched for only one MHC class II antigen for assays of R and Ci clones, or were fully mismatched (incompat.) or not tested (NT), in the presence of the 25% of the various clones¹⁶.

To confirm the expression of TCR α and β on R9 cells, mRNA expression was determined by Northern blot analysis. Full length TCR α and both full length (1.3 kilobase pairs (kb)) and short (1.0 kb) TCR β mRNAs were detected in Northern blot analyses using ³²P-labelled TCR α and β cDNA probes, respectively (Fig. 2a). The presence of full length TCR α and β mRNA offers further evidence for expression of a complete TCR $\alpha\beta$ heterodimer.

Previous studies on murine T suppressor hybridomas indicated that TCR β genes were either germline or deleted¹¹⁻¹³. Since TCR β mRNA and polypeptides were detected on the T_S clones examined here, it seemed that these lymphocytes would necessarily have rearranged their TCR β genes. To verify this, Southern blot analyses were performed on high molecular weight R9 and Ci1 DNA digested with *Hind*III and *Eco*RI, using ³²P-labelled J β ₁ and J β ₂ genomic probes, respectively²⁸. Using these enzyme-probe combinations, J β ₁ rearrangements were detected in both R9 and Ci1 DNA, while J β ₂ rearrangements were detected only in R9 DNA. Thus, T_S clones were demonstrated to rearrange their TCR β genes (Fig. 2b). The clonality of R9 and Ci1 was supported by the demonstration that two or fewer rearranged DNA fragments from each sample hybridized to the J β ₁ or J β ₂ probes.

The expression of TCR γ mRNA in T_S cells is of particular interest, as γ mRNA is expressed at high levels in neonatal thymus^{33,34} (when self-tolerance is presumably being induced) and in-frame, functional TCR γ mRNA has generally not been found in T_H or T_C clones³⁵⁻³⁸. TCR γ mRNA was undetectable in R9 cells (Fig. 2a), suggesting that a cell-surface TCR γ polypeptide would not be displayed by this T_S clone. To rule out the possible co-expression of a TCR γ polypeptide with TCR $\alpha\beta$ polypeptides on the T_S clones, TCR γ -specific antisera were used. Two such antisera have been shown previously to recognize the cell surface protein product of the rearranging TCR γ gene. (Anti-C γ _a, previously reported as anti-C γ ₂₄, reacts with all human TCR γ cells examined thus far, anti-C γ _{2c} reacts with a subset of TCR γ lymphocytes (M.B.B. unpublished data) and reactivity of anti-V γ _b is under study.) Accordingly, aliquots of

R9, R2, Ci1, J301 and J303 radiolabelled cell lysates were denatured, then partially renatured and subjected to immunoprecipitation with these antisera generated against synthetic peptides corresponding to three different 15-20 amino-acid residue segments of a deduced TCR γ amino-acid sequence^{35,39}. None of these anti-TCR γ peptide reagents immunoprecipitated a TCR γ protein from the surfaces of these T_S cells (Fig. 1a, b, c, lanes 8-10 for each clone; and Fig. 1d, e, lane 6 for each clone).

Taken together: (1) the virtual absence of TCR γ mRNA; (2) the absence of reactivity with anti-C γ sera; (3) the demonstration of TCR β gene rearrangements; (4) the expression of full length TCR α and β mRNA; (5) the successful isolation of CD3-co-immunoprecipitated, heterodimeric acidic and basic subunits; (6) the isolation of these same subunits using the TCR β chain reactive β F1 monoclonal antibody argue compellingly that the T_S cells examined here express TCR α and β , but not γ proteins.

It could be argued that the culture conditions used to propagate the leprosy skin lesion-derived T_S clones may have selected for a minor subset of CD8⁺ cells. To determine whether the receptors found on these T_S clones are representative of the majority of cells infiltrating lepromatous lesions *in vivo*, we carried out immunoperoxidase staining and microscopic analysis of the lymphocytes present *in situ* in freshly isolated skin biopsies⁴⁰. In specimens from 14 lepromatous leprosy patients, 43% of infiltrating cells were CD3⁺, 23% were CD8⁺ and 30% were β F1⁺. As it was conceivable that a significant portion of the CD8⁺ cells were β F1⁻, we investigated the phenotype of the CD8⁺ cells in specimens from five lepromatous leprosy patients with double immunoperoxidase-glucose oxidase staining. The majority (>80%) of primary CD8⁺ cells in lepromatous lesions expressed the TCR $\alpha\beta$ complex detected by mAb β F1 (Fig. 3). The possibility that some of the remaining *in situ* CD8⁺ cells might express the TCR $\gamma\delta$ protein complex has not been excluded.

In summary, while there may well be more than one cell type or immune mechanism that mediates suppression, the *in vitro* and *in vivo* studies presented here establish that TCR $\alpha\beta$ complexes are expressed on one subset of human, antigen-responsive suppressor T cells.

We are grateful to Dr T. H. Rea for classification and staging of the Hansen's disease patients, and for support and encouragement, and to E. Nelson, L. Chen, G. Gersuk and Dr J. Melancon-Kaplan for help and advice in preparing the CD8⁺ cells. We thank Drs V. Mehra, F. Hofman and A. Rao for discussions, Drs P. Beverley, W. Tax, J. G. Seidman and S. Ip for antibodies, N. Dewese for performing the immunoperoxidase staining and Dr V. Wong for cytofluorographic analysis. Supported by the NIH, ACS, UNDP/World Bank/WHO Special Programme for Research and Training in Tropical Diseases (IMMLEP) and Special Programme for Vaccine Development (IMMTUB), the Hartford Foundation, the Medical Research Council of Canada, and the Irvington House Institute for Medical Research.

Received 14 July; accepted 14 August 1987.

- Green, D. R., Flood, R. M. & Gershon, R. K. *A. Rev. Immun.* **11**, 439-464 (1983).
- Dorf, M. E. & Benacerraf, B. *A. Rev. Immun.* **2**, 127-158 (1984).
- Yague, J. *et al. Cell* **42**, 81-87 (1985).
- Dembic, Z. *et al. Nature* **320**, 232-238 (1986).
- Saito, T., Weiss, A., Miller, J., Norcross, M. A. & Germain, R. N. *Nature* **325**, 125-130 (1987).
- Yoshikai, Y., Yanagi, Y., Suciu-Foca, N. & Msk, T. W. *Nature* **310**, 506-508 (1984).
- Royer, H. D., Bensussan, A., Acuto, O. & Rheiner, E. L. *J. exp. Med.* **160**, 947-952 (1984).
- Toyonaga, B., Yanagi, Y., Suciu-Foca, N., Minden, M. & Mak, T. W. *Nature* **311**, 385-2387 (1984).
- Ballinari, D. *et al. Eur. J. Immun.* **15**, 855-860 (1985).
- De Santis, R. *et al. Proc. natn. Acad. Sci. U.S.A.* **82**, 8638-8642 (1985).
- Hedrick, S. M. *et al. Proc. natn. Acad. Sci. U.S.A.* **82**, 531-535 (1985).
- Kronenberg, M. *et al. Nature* **313**, 647-653 (1985).
- Blankmeister, C. A., Yamamoto, K., Davis, M. M. & Hammerling, G. J. *J. exp. Med.* **162**, 751-863 (1985).
- Mori, L. *et al. EMBO J.* **4**, 2025-2030 (1985).
- Modlin, R. L. *et al. J. Immun.* **137**, 2831-2834 (1986).
- Modlin, R. L. *et al. Nature* **322**, 459-461 (1986).
- Bloom, B. R. & Godal, T. *Rev. infect. Dis.* **5**, 765-779 (1983).
- Bloom, B. R. *J. Immun.* **137**, i-x (1986).

19. Mehra, V., Mason, L. H., Fields, J. P. & Bloom, B. R. *J. Immun.* **123**, 1813-1817 (1979).
20. Mehra, V., Convit, J., Rubinstein, A. & Bloom, B. R. *J. Immun.* **129**, 1946-1951 (1982).
21. Nelson, E., Wong, L., Uyamura, K., Rea, T. H. & Modlin, R. L. *Cell. Immun.* **104**, 99-104 (1987).
22. Oettgen, H. C., Kappler, J., Tax, W. J. M. & Terhorst, C. *J. biol. Chem.* **259**, 12039-12048 (1984).
23. Spits, H. *et al. J. Immun.* **135**, 1922-1928 (1985).
24. Brenner, M. B. *et al. Nature* **322**, 145-149 (1986).
25. Brenner, M. B. *et al. J. Immun.* **138**, 1502-1509 (1987).
26. Brenner, M. B. *et al. Nature* **325**, 689-694 (1987).
27. De Vries, R. R. P., Ottenhoff, T. H. M., Li, S. & Young, R. A. *Lepr. Rev.* **57**, suppl. 2, 113-121 (1986).
28. Borst, J., Prendville, M. A. & Terhorst, C. *J. Immun.* **128**, 1560-1565 (1982).
29. Mauer, S. C. *et al. J. exp. Med.* **157**, 705-719 (1983).
30. Reinherz, E. L. *et al. Proc. natn. Acad. Sci. U.S.A.* **80**, 4104-4108 (1983).
31. Weiss, A. & Stobo, J. D. *J. exp. Med.* **160**, 1284-1299 (1984).
32. Brenner, M. B., Trowbridge, I. S. & Strominger, J. L. *Cell* **40**, 183-190 (1985).
33. Raulet, D. H., Garman, R. D., Saito, H. & Tonegawa, S. *Nature* **314**, 103-107 (1985).
34. Snodgrass, H. R., Dembic, Z., Steinmetz, M. & von Boehmer, H. *Nature* **315**, 232-233 (1985).
35. Dialynas, D. P. *et al. Proc. natn. Acad. Sci. U.S.A.* **83**, 2619-2623 (1986).
36. Quertermous, T. *et al. Nature* **322**, 184-187 (1986).
37. Rupp, F., Frech, G., Hengartner, H., Zinkernagel, R. M. & Joho, R. *Nature* **321**, 876-878 (1986).
38. Reilly, E. B., Kranz, D. M., Tonegawa, S. & Eisen, H. N. *Nature* **321**, 878-880 (1986).
39. Murre, C. *et al. Nature* **316**, 549-552 (1985).
40. Modlin, R. L. *et al. J. Immun.* **132**, 3085-3090 (1984).
41. Ottenhof, T. H. M., Elferink, D. G., Klatzer, P. R. & de Vries, R. R. P. *Nature* **322**, 462-464 (1986).
42. Fox, E. J., Cook, R. G., Lewis, D. E. & Rich, R. R. *J. clin. Invest.* **78**, 214-220 (1986).
43. Brenner, M. B., Trowbridge, I. S., McLean, J. & Strominger, J. L. *J. exp. Med.* **160**, 541-551 (1984).
44. Samelson, L. E., Harford, J., Schwartz, R. H. & Klausner, R. D. *Proc. natn. Acad. Sci. U.S.A.* **82**, 1969-1973 (1985).
45. Yelton, D. E., Desaymard, C. & Scharff, M. D. *Hybridoma* **1**, 5-11 (1981).
46. Yanagi, Y. *et al. Nature* **308**, 145-149 (1984).
47. Wigler, M. *et al. Cell* **16**, 778-785 (1979).
48. Toyonaga, B., Yoshikai, Y., Vadasz, V., Chin, B. & Mak, T. W. *Proc. natn. Acad. Sci. U.S.A.* **82**, 8624-8628 (1985).
49. Duby, A. D., Klein, K. A., Murre, C. & Seidman, J. G. *Science* **228**, 1204-1206 (1985).

Recovery of immunodeficient mice from a vaccinia virus/IL-2 recombinant infection

Ian A. Ramshaw*, Marion E. Andrew†‡, Susan M. Phillips*, David B. Boyle†‡ & Barbara E. H. Coupard†‡

Departments of † Microbiology and * Experimental Pathology, John Curtin School of Medical Research, Australian National University, GPO Box 334 Canberra, ACT 2601, Australia

Vaccinia virus recombinants that express cloned genes encoding antigens of unrelated infectious agents, such as hepatitis B virus¹ and human immunodeficiency virus (HIV)^{2,3}, provide a new approach to the development of live vaccines. Although there is evidence that genetically engineered vaccinia viruses have reduced pathogenicity⁴ a major obstacle to their use as vaccines is that severe complications can occur after vaccination, especially in immunodeficient individuals^{5,6}. We describe here a recombinant vaccinia virus expressing murine interleukin-2 (IL-2) and show that athymic nude mice infected with the recombinant virus resolve the virus infection rapidly whereas mice infected with control virus develop a progressive vaccinal disease. By incorporating the gene for IL-2 in live virus vaccines it may be possible to prevent the severe complications that arise in recipients with an impaired immune system.

IL-2 is an immunoregulatory, T-cell-derived molecule which is required for the clonal expansion of antigen-activated T cells⁷. A recombinant vaccinia virus expressing murine IL-2 was constructed to study the lymphokine's effect on virus growth and immunogenicity. As shown schematically in Fig. 2a, complementary DNA encoding murine IL-2 (ref. 8) was inserted into the *Hind*III F region of a vaccinia recombinant, VV-HA (ref. 9), which expresses the influenza haemagglutinin. The IL-2 recombinant virus, VV-HA-IL2, coexpressed haemagglutinin

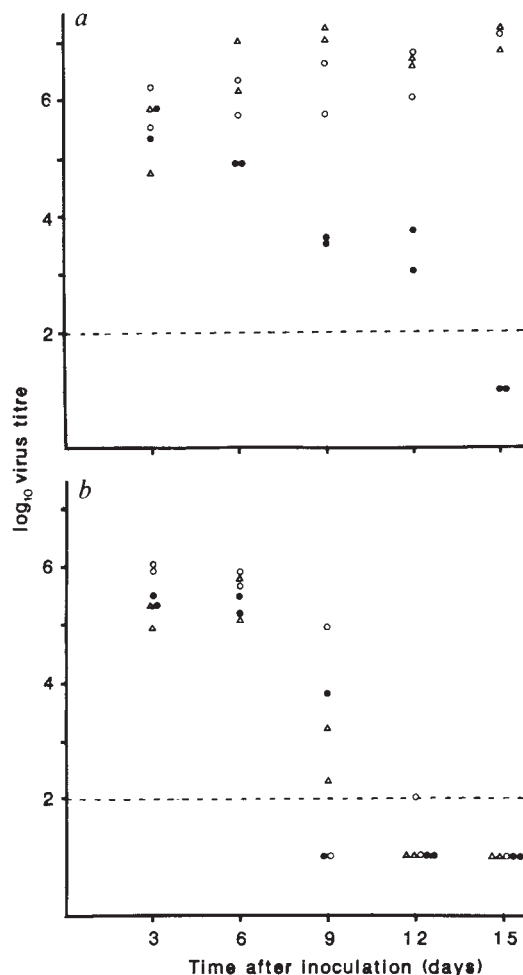


Fig. 1 Growth of vaccinia virus recombinants in the foot pads of athymic Swiss outbred nude mice (a) and euthymic CBA/H mice (b). VV-HA (Δ), VV-HA-TK ($\circ$) or VV-HA-IL2 ($\bullet$) (2×10^7 PFU of each in 20 μ l of saline) were injected subcutaneously into hind foot pads which were assayed for infectious virus on 143B cells on the indicated days. Points represent the titres of infectious virus present in individual mice.

and IL-2 using the same vaccinia 7.5-kD promoter but from separate sites in the viral genome. Because the herpes simplex virus (HSV) thymidine kinase (TK) gene was used as a selectable marker for virus construction, a control virus VV-HA-TK, expressing HSV TK but not IL-2 was constructed. A significant amount of biologically active IL-2 was detected in supernatants from human 143B cells infected with VV-HA-IL2 within 4 h and reached maximum activity around 12 h after infection (Fig. 2b).

Athymic nude mice (4-8 weeks old) were inoculated into the right hind footpad with VV-HA-IL2 or control virus (VV-HA-TK). VV-HA-IL2 induced a mild swelling in the foot which resolved after several days; in contrast VV-HA-TK produced a severe necrotic lesion that remained unresolved for 30 days. After this time, high titres of virus (6×10^5 to 1.5×10^7 plaque-forming units (PFU)) were recovered from the feet of the VV-HA-TK inoculated mice but not from mice given VV-HA-IL2. This suggested that the IL-2 produced by the recombinant virus enabled immunodeficient mice to control the virus infection. The kinetics of viral clearance from the feet of CBA/H mice were not significantly different for VV-HA-TK and VV-HA-IL2 (Fig. 1b). In nude mice, however, although titres of both VV-HA-TK and VV-HA-IL2 were high at day 3, indicating comparable rates of replication, VV-HA-IL2 was cleared by day 15, when no virus was detected in the feet. Titres of VV-HA-TK still remained high at day 15 (Fig. 1a). Furthermore, when nude

‡ Permanent address: CSIRO Australian Animal Health Laboratory, Geelong, Australia.

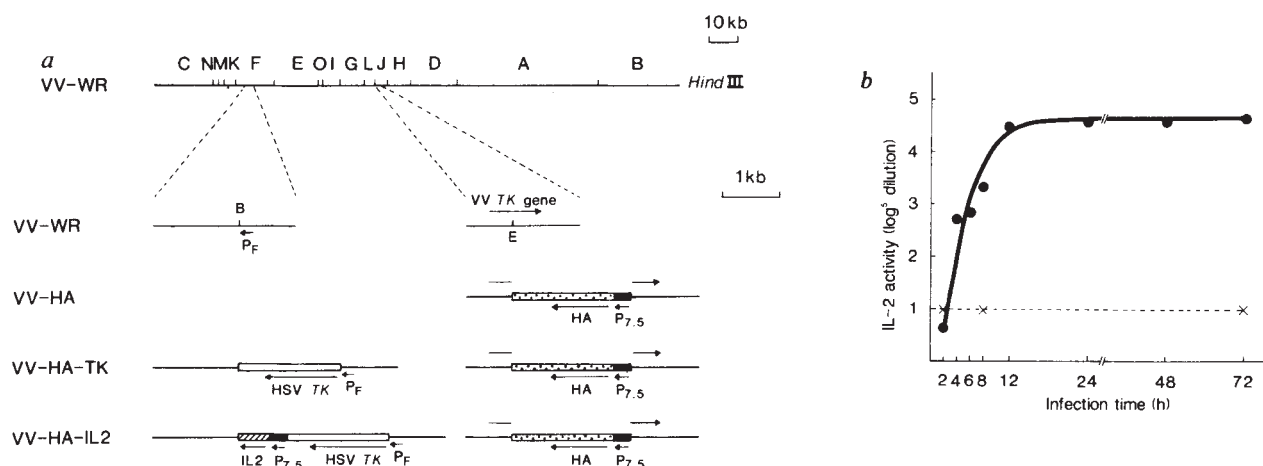


Fig. 2 *a*, Genomic configuration of vaccinia virus recombinants. A *Hind*III map of vaccinia virus strain WR (VV-WR) is shown with insertion points at the *Eco*RI (*E*) and *Bam*HI (*B*) sites in the J and F fragments respectively. Arrows indicate orientations of vaccinia virus TK gene, vaccinia promoters and inserted influenza haemagglutinin, HSV TK and murine IL-2 coding sequences. *b*, Time course of IL-2 production by VV-HA-IL2-infected human 143B cells. ●, IL-2 activity in VV-HA-IL2-infected supernatants, (solid line); ×, VV-HA or uninfected cell supernatants (dotted line).

Methods. Murine IL-2 cDNA was subcloned from pcD-IL-2 (ref. 8) (K. I. Arai, DNAX Research Institute, Palo Alto, California). VV-HA-IL2 and VV-HA-TK were constructed by insertion of the HSV TK gene plus a chimaeric promoter-IL-2 fragment or alone, into the *Hind*III F region of the recombinant vaccinia virus haemagglutinin (described in ref. 9 as VV-PR8-HA6) using plasmids and selection protocols to be described elsewhere (B.E.H.C., M.E.A. and D.B.B., in preparation). *b*, Human 143B cells (2×10^6) were infected with VV-HA-IL2 at 5 PFU cell $^{-1}$. The supernatants (1.5 ml), collected at the time points indicated, were assayed for IL-2 activity using CTLL-2 cells 12 and the colorimetric method for cell growth of Mosmann 13 . The results are presented as the log $_2$ of supernatant producing 50% of maximum proliferation in the cell cultures.

mice were injected intravenously (i.v.) with 10^6 PFU of VV-HA-IL2 or VV-HA-TK, mice given VV-HA-IL2 seemed unaffected by the virus whereas all mice given VV-HA-TK were dying by day 15. Consistent with this result infectious virus was recovered from spleens and lungs of mice infected with VV-HA-TK but not those infected with VV-HA-IL2 (Table 1).

We have not studied in depth the mechanisms of this recovery, although in nude and normal CBA/H mice, VV-HA-IL2 induced significantly smaller amounts of antibody than VV-HA-TK, suggesting that recovery was not antibody mediated. The kinetics of virus clearance (Fig. 1) suggest that cell-mediated immunity is involved and we are currently investigating possible immune effector mechanisms including cytotoxic T cells, NK cells and lymphokine-activated killers.

Our results have several implications. Live virus vaccines that express IL-2 may minimize or eliminate the complications associated with the inadvertent vaccination of immunodeficient recipients. This may be particularly important if vaccines are to be used in countries where HIV is endemic but prior screening is impracticable. The development of generalized vaccinia and subsequent death from AIDS (acquired immune deficiency syndrome) of an army recruit with subclinical HIV-1 infection, vaccinated for smallpox illustrates the potential risks involved 6 . Inclusion of an IL-2 gene may make vaccination with vaccinia

virus recombinants possible for a variety of immunocompromised individuals, such as those with congenital immunodeficiencies, or transplant patients on immunosuppressive therapy. Other live recombinant virus vaccines (herpes or adenoviruses) expressing IL-2 may also prove valuable.

The most intriguing possibility is that VV-IL-2 recombinants could be used for the treatment of AIDS patients. Some of the functional abnormalities in the immunopathogenesis of AIDS have been associated with deficiency in IL-2 production 10 . Recombinant vaccinia viruses expressing the HIV *env* gene have been constructed, but their use in AIDS patients to boost or modify immune response would be contra-indicated. Concomitant expression of IL-2 by the recombinant virus might assist in promoting an appropriate immune response in these patients. This must be balanced, however, against potential acceleration of the disease by lymphocyte activation 11 .

Finally, IL-2 is only one of many lymphokines involved in immune regulation and we are currently studying the effects of other lymphokines expressing in vaccinia virus on the immune responses to viral antigens.

Technical assistance was provided by Ms A. Kasprzak, Mr C. Rolls, Ms J. Thistleton and Mrs M. Maxted. The 7.5Kd VV promoter was provided by Dr B. Moss, NIH.

Received 17 July; accepted 1 September 1987.

- Smith, G. L., Mackett, M. & Moss, B. *Nature* **302**, 490-495 (1983).
- Chakrabarti, S., Robert-Guroff, M., Wong-Staal, F., Gallo, R. C. & Moss, B. *Nature* **320**, 535-537 (1986).
- Hu, S.-L., Kosowski, S. G. & Dalrymple, J. M. *Nature* **320**, 537-540 (1986).
- Buller, R. M. L., Smith, G. L., Cremer, K., Notkins, A. L. & Moss, B. *Nature* **317**, 813-815 (1985).
- Dick, G. *Prog. med. Virol.* **8**, 1-29 (1966).
- Redfield, R. R. *et al. N. Engl. J. Med.* **316**, 673-676 (1987).
- Smith, K. A. *A. Rev. Immun.* **2**, 319-333 (1984).
- Yokota, T. *et al. Proc. natn. Acad. Sci. U.S.A.* **82**, 68-72 (1985).
- Andrew, M. E., Coupar, B. E. H., Ada, G. L. & Boyle, D. B. *Microbial Pathogenesis* **1**, 443-452 (1986).
- Lifson, J. D., Benike, C. J., Mark, D. F., Koths, K. & Engleman, E. G. *Lancet* **i**, 698-702 (1984).
- Gluckman, J. C., Klatzmann, D. & Montagnier, L. *A. Rev. Immun.* **4**, 97-117 (1986).
- Gillis, S., Fehm, M. M., Ou, W. & Smith, K. A. *J. Immun.* **120**, 2027-2032 (1978).
- Mosmann, T. *J. Immunol. Meth.* **65**, 55-63 (1983).

Table 1 Vaccinia virus recovered from lungs and spleens

Mouse strain	Organ	log $_{10}$ VV-HA-TK	log $_{10}$ VV-HA-IL2
CBA/H	Lung	<2.0	<2.0
	Spleen	<2.0	<2.0
Outbred nude	Lung	4.91 $\pm$ 0.27	<2.0
	Spleen	3.64 $\pm$ 0.11	<2.0

Lungs and spleens are collected from 3-4 mice in each group 11 days after i.v. inoculation with 10^6 PFU of the indicated virus.

Antibody production to the nucleocapsid and envelope of the hepatitis B virus primed by a single synthetic T cell site

David R. Millich†, Alan McLachlan,
George B. Thornton* & Janice L. Hughes

Department of Molecular Biology, Scripps Clinic
and Research Foundation, 10666 North Torrey Pines Road,
La Jolla, California 92037, USA

* Biotechnology Center, Inc., San Diego, California 92038, USA

The nucleocapsid (HBcAg) of the hepatitis B virus (HBV) can induce antibody responses via both a T-cell dependent and a T-cell independent pathway¹ and is highly immunogenic during infection. We have examined the T-cell determinants of the antigen and find that HBcAg-specific helper T cells (T_H) can help B cells produce antibody against envelope (HBsAg) antigens as well as HBcAg, even though these antigens are found on separate molecules. We have also been able to prime helper T cells with synthetic T-cell epitopes of HBcAg; helper cells primed with a single synthetic epitope can induce B cells to produce antibody that reacts with multiple HBsAg epitopes. One problem with the development of an HBV vaccine is that some vaccinees and patients do not respond to HBsAg directly; our results indicate that this problem can be circumvented using the response to HBcAg.

The hepatitis B virus (HBV) consists of an outer envelope of three co-terminal polypeptides (p25, gp33, p39) and an inner nucleocapsid, which encapsidates the viral genome. Envelope and nucleocapsid-specific cellular immune responses have been suggested to be important in virus elimination and the attendant hepatocellular injury^{2,3}, and vaccination with both antigens has been reported to protect against HBV infection⁴⁻⁶. As antibodies to the HBcAg do not neutralize the virus, the mechanism of protection by HBcAg is unknown.

The hepatitis B surface envelope (HBsAg) is composed of a major polypeptide, p25⁷. The larger polypeptides share the 226 amino acids of p25 (S region) at the C terminus and possess additional residues at the N terminus. The glycoprotein (gp)33 consists of the p25 sequence plus an N-terminal 55 residues (pre-S(2) region), and p39 consists of the gp33 sequence plus an N-terminal 108 or 119 residues (pre-S(1) region)^{8,9}. The nucleocapsid of the HBV is a 27-nm particle composed of multiple copies of a single polypeptide (p21), which exhibits

hepatitis B core antigenicity (HBcAg). Treatment of HBcAg with nonionic detergent or protease destroys HBcAg antigenicity and cryptic HBcAg determinants are exposed.

Multiple, but distinct, T-cell sites on HBcAg are recognized by individual inbred murine strains¹⁰. The site recognized is dependent on the H-2 haplotype of the responding strain. For example, H-2^{s,b} strains recognize p120-140, H-2^{f,a} strains recognize p100-120, and H-2^d mice recognize p85-100, predominantly¹⁰. Preliminary data indicate that those strains which appear to 'see' the same T-cell site in fact recognize different residues within the larger peptide. This murine model predicts that a human outbred population would exhibit similar complexity, and individuals may recognize distinct T-cell sites in the context of their HLA genotype. Two synthetic peptides, p85-100 and p120-140, were chosen to examine the functional characteristics of HBcAg-primed T cells in Balb/c (H-2^d) and B10.S (H-2^s) mice, respectively. Before examining the ability of the synthetic peptides to prime T_H cell activity for anti-HBc production, it was necessary to determine if immunization with the synthetic T-cell epitopes would elicit anti-peptide antibody crossreactive with native HBcAg. Antibody to HBcAg is not reactive with p120-140 or p85-100. Immunization with p120-140 or p85-100 (100 µg) induced antibody specific for the immunizing peptide, but nonreactive with HBcAg. Thus, although the peptide sequences possess HBcAg-specific, T-cell determinants, the B-cell epitopes present on the peptides are either not relevant to or not exposed on the native HBcAg.

Therefore, it was possible to examine the ability of these peptides to prime T_H cells *in vivo* directly, as opposed to performing T-cell transfer experiments. B10.S mice were primed with p120-140 or p85-100, and three weeks later challenged with a suboptimal dose of either native HBcAg or HBV in incomplete adjuvant. Seven days after challenge, sera were analysed for IgG, anti-HBc antibody. This approach requires that the memory T cells primed by immunization with peptide be recalled by challenge with native HBcAg, indicating that the synthetic T-cell recognition site is relevant to the native molecule. Unprimed B10.S mice did not produce anti-HBc after challenge with HBcAg or HBV, nor did p120-140-primed and unchallenged mice produce anti-HBc (Fig. 1). However, p120-140-priming elicited high titred, anti-HBc production after challenge with HBcAg and HBV. The p85-100 sequence did not prime anti-HBc production in B10.S mice; however, Balb/c mice were primed by this sequence as high titred, anti-HBc was produced after challenge with HBcAg (1:8192) (Fig. 1). Thus HBcAg-specific synthetic, T-cell determinants can prime T cells which function as helper T cells to elicit anti-HBc antibody *in vivo*.

† To whom correspondence should be addressed.

Table 1 Fine specificity of anti-envelope antibody production after priming with HBcAg or p120-140 in B10.S mice

Prime*	Challenge	Time (d)	S region		Antibody titre (1 dilution ⁻¹)				
			Subtype†	Group	Pre-S(2) region‡		Pre-S(1) region§		
0	HBV	7	0	0	0	0	0	0	0
		14	0	0	0	0	0	0	0
HBcAg	HBV	7	640	0	80	40	160	80	320
		14	2,560	160	320	640	1,280	2,560	1,280
p120-140	HBV	7	80	0	160	80	160	0	160
		14	320	40	1,280	160	320	320	160

* Groups of four B10.S mice were primed with CFA alone (0), 4.0 µg of HBcAg in CFA, or 100 µg of p120-140 in complete Freund's adjuvant (CFA). Three weeks later, all mice were challenged with a suboptimal dose of HBV (0.4 µg) in incomplete adjuvant, and sera were tested by radioimmunoassay for IgG antibody against the panel of antigens 7 and 14 days after the challenge. Titres are expressed as the reciprocal of the serum dilution which yielded four times the counts of preimmunization sera.

† HBsAg/p25 of the *ad* subtype was the solid-phase ligand used to detect subtype plus group-specific anti-S, whereas, HBsAg/p25 of the *ay* subtype was the solid-phase ligand used to detect group-specific anti-S.

‡ The pre-S(2) region synthetic peptides used have been previously shown to represent the dominant antibody binding sites within the pre-S(2) region²².

§ The pre-S(1) region synthetic peptides used have been previously shown to represent the dominant antibody binding sites within the pre-S(1) region^{14,15}.

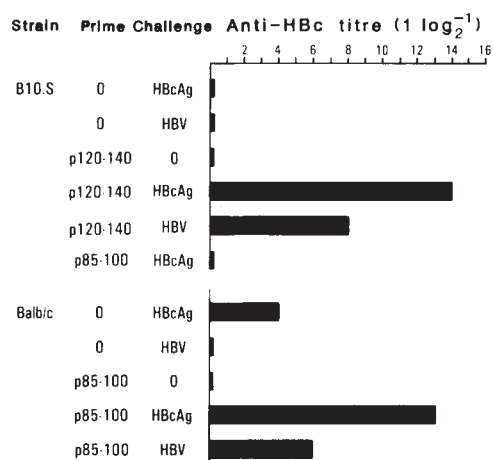


Fig. 1 Synthetic T-cell determinants of HBcAg can prime T_H cells which induce anti-HBc production *in vivo*. Groups of five B10.S (upper panel) or Balb/c (lower panel) mice were primed by intraperitoneal (i.p.) immunization with 100 μ g of synthetic peptide (p120-140 or p85-100) in CFA or CFA alone (0). After three weeks, the primed mice were challenged with either a suboptimal dose of recombinant HBcAg (Biogen, SA) (0.1 μ g) or HBV (0.4 μ g) in incomplete adjuvant or with adjuvant alone (0). Seven days after the challenge dose, sera were collected, pooled and analysed for IgG, anti-HBc antibody by solid-phase radioimmunoassay (RIA) as described previously¹. The anti-HBc titre is expressed as the reciprocal of the log₂ of the highest serum dilution required to yield four times the counts of sera before immunization. The peptides were synthesized by the Merrifield solid-phase method, and were subjected to HPLC on a C18 reverse phase column. All peptides used eluted as a single major peak (>90%).

Classic cognate T-cell help requires direct T-cell-B-cell interaction mediated by an antigen bridge. Therefore, the antigenic determinants recognized by T_H and B cells must be covalently linked on the same molecule¹¹. However, different rules may apply to particulate antigens such as viruses. A multimolecular particle is bound and internalized by a specific B cell, and all T-cell sites on all molecules within the particle may be re-expressed on the B-cell surface; each may serve as ligand for T_H -cell recognition. Such a mechanism of T-cell help has been observed in the influenza system¹², and has been called intermolecular-intrastructural¹³. In view of this, we examined the ability of HBcAg-primed T cells to function as T_H cells for antibody production to envelope (HBsAg) epitopes even though HBsAg and HBcAg are on distinct molecules, albeit within the same virion. B10.S mice were primed with either HBcAg or complete Freund's adjuvant (CFA) alone, and challenged three weeks later with a suboptimal dose of either a mixture of HBcAg and HBsAg/p39 (possesses pre-S(1), pre-S(2) and S region antigens), or HBV. Unprimed mice produced no envelope-specific antibody (Fig. 2). Similarly, mice primed with HBcAg and challenged with a mixture of HBcAg and HBsAg/p39 produced no anti-HBs. However, mice primed with HBcAg and subsequently challenged with virions (HBV) produced anti-S, anti-pre-S(2), and anti-pre-S(1)-specific antibodies (Fig. 2). This result indicates that HBcAg-primed T cells can function to help anti-envelope antibody production, and the T_H -cell activity required that HBcAg and HBsAg be present within the same particle (virion), as the mixture was ineffective.

To confirm the T-cell nature of this effect, an identical experiment was performed using the synthetic T-cell recognition site, p120-140, as the priming antigen. The results obtained were similar to those using native HBcAg as the priming antigen (Fig. 2). Mice primed with p120-140 and challenged with HBV produced anti-envelope antibodies, whereas p120-140-primed mice challenged with the mixture did not. The B10.S strain is nonresponsive to the S region of HBsAg, yet priming with a HBcAg-specific peptide circumvented nonresponse to

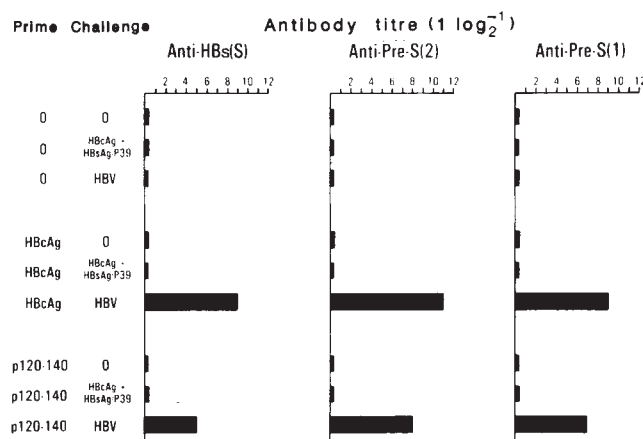


Fig. 2 HBcAg-primed and p120-140-primed T_H cells of B10.S mice can induce antibody production specific for the S, pre-S(2) and pre-S(1) regions of the envelope of the HBV. Groups of five B10.S mice were primed by i.p. immunization with either CFA alone (0, upper panel), 4.0 μ g of HBcAg in CFA (middle panel), or 100 μ g of the synthetic peptide p120-140 in CFA (lower panel). Three weeks after priming, mice were challenged either with incomplete adjuvant alone (0), or a suboptimal dose of a mixture of HBcAg (0.1 μ g) and HBsAg/p39 (0.6 μ g), or with HBV (0.4 μ g) in incomplete adjuvant. The HBsAg/p39 and the HBV preparations (provided by J. Gerin, Georgetown University, Rockville, Maryland, USA) were of the *adw2* subtype purified from the plasma of a chronic HBsAg carrier by biophysical methods²³, and were equilibrated to contain approximately equivalent amounts of S, pre-S(2) and pre-S(1) region antigenicity as described previously¹⁵. Seven days after the challenge dose, sera were collected, pooled and analysed for IgG antibody specific for the S, pre-S(2), and pre-S(1) regions of HBsAg by solid-phase RIA¹⁵. The solid-phase ligands used to detect anti-S, anti-pre-S(2) and anti-pre-S(1)-specific antibodies were: HBsAg/p25, a pre-S(2) region sequence represented by p133-145 (ref. 22), and a pre-S(1) region sequence represented by p94-117 (ref. 15), respectively. The antibody titres are expressed as the reciprocal of the log₂ of the highest serum dilution required to yield four times the counts of sera before immunization.

HBsAg/p25 by providing an alternate T_H cell recognition site on the HBV which B10.S strain T cells recognized. These results show that intermolecular-intrastructural T_H -cell function can be mediated by a synthetic T-cell epitope.

Next we examined the diversity of anti-envelope antibodies produced under the influence of HBcAg and p120-140-primed T_H cells. As shown in Table 1, priming with HBcAg or the synthetic peptide p120-140 resulted in antibody production to seven distinct epitopes within the S and pre-S regions of HBsAg after challenge with HBV. Therefore, HBcAg-primed T cells are competent to provide functional T-cell help to multiple B-cell clones specific for a variety of antigenic determinants of the envelope, including three pre-S(1) epitopes, two pre-S(2) epitopes and group and subtype-specific epitopes of the S region.

Regulation of the immune responses to the pre-S and S regions of HBsAg has been extensively studied. T_H cell recognition of a single site within one region is sufficient to induce antibody production to multiple S and pre-S region epitopes¹⁴. Therefore, S and pre-S(2) region nonresponsiveness can be circumvented by immunization with HBsAg/p39 (Fig. 3) (refs 15-17). Cumulatively, these observations suggest that, after vaccination with HBsAg/p39, all or none of the anti-envelope specificities should be produced. In general, the clinical observations also support this prediction¹⁸. However, a subset of chronically infected HBV patients has been reported to produce only anti-pre-S(1) in the absence of anti-pre-S(2) or anti-S antibodies, which correlated with seroconversion from HBeAg to anti-HBe status and viral clearance¹⁹. This observation suggests that additional factors may influence antibody production during natural infection as opposed to vaccination.

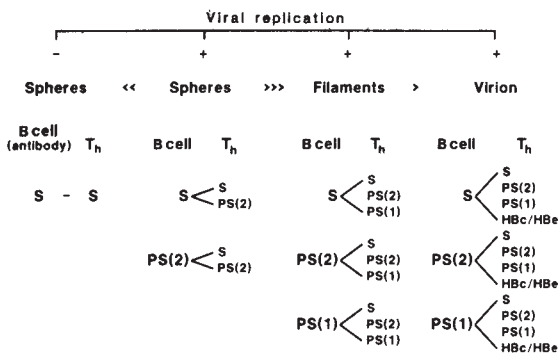


Fig. 3 Schematic representation of T helper cell-B cell interactions regulating the humoral immune responses to the envelope antigens of the hepatitis B virus and subviral particles (spheres, filaments). The B-cell (antibody) columns depict antigenic specificities to which antibodies are produced. The T_H-cell columns depict the specificities of T_H cells which can provide functional help for antibody production to the various B-cell determinants. All the T_H-B cell interactions shown have been experimentally substantiated¹⁴⁻¹⁷, with the exception of the ability of S-specific T_H cells to help anti-pre-S antibody production, which is inferred. The relative quantities of virions and subviral particles circulating in the serum, and the differential expression of the pre-S antigens on the various morphological forms during and in the absence of viral replication are also depicted (see text).

We propose that anti-pre-S(1) production in the absence of anti-pre-S(2) and anti-S production is mediated by T_H cells specific for HBcAg/HBeAg determinants. This is consistent with the correlation of anti-pre-S(1) production with anti-HBe seroconversion¹⁹, and the ability of HBcAg/HBeAg-specific T cells to help envelope-specific antibody production as described here. The question of why only anti-pre-S(1) is produced may be explained by the fact that subviral, 22-nm spherical particles lacking nucleocapsid antigens circulate at levels in great excess of intact virions during HBV infection. The subviral particles express both S and pre-S(2) region antigens but no or only minimal pre-S(1) region antigens (Fig. 3). Therefore, B cells specific for S and pre-S(2) region determinants are far more likely to bind subviral particles than virus, and will not benefit from HBcAg-specific T_H cells. Overproduction of subviral particles may represent a viral strategy to evade this aspect of the host's immune response. In contrast, pre-S(1)-specific B cells are more likely to recognize their ligand on intact virions as p39 is preferentially associated with the virus⁸.

The fact that HBcAg-specific T_H cells can elicit anti-envelope antibodies, which are virus neutralizing^{4,20}, may explain the reported ability of HBcAg vaccination to protect against HBV infection^{5,6}. Furthermore, HBcAg-specific T_H cells were shown to induce anti-S antibody production in S region nonresponder mice. This report describes a mechanism of circumventing HBsAg nonresponsiveness in which the only requirement was that HBcAg and HBsAg be present within the same particle (for example, HBV). This observation can be applied to vaccine development. The recognized immunogenicity of HBcAg in humans during HBV infection, and the characterization of the underlying cellular mechanisms^{1,10} suggest that HBcAg may serve as an excellent carrier vehicle for a number of human and veterinary vaccines in addition to HBV, which require T_H cell augmentation. It has been proposed that HBsAg be used as a particulate, carrier moiety for multivalent vaccines²¹; however, HBcAg is approximately 100-fold more efficient than HBsAg in terms of both its ability to elicit antibody production¹ and activate T cells¹⁰.

5. Tabor, E. & Gerety, R. J. *Lancet* **i**, 172-173 (1984).
6. Murray, K. *et al. EMBO J.* **3**, 645-650 (1984).
7. Peterson, D. L., Roberts, I. M. & Vyas, G. N. *Proc. natn. Acad. Sci. U.S.A.* **74**, 1530-1537 (1977).
8. Heerman, K. H. *et al. J. Virol.* **52**, 396-403 (1984).
9. Machida, A. *et al. Gastroenterology* **86**, 910-917 (1984).
10. Milich, D. R., McLachlan, A., Moriarty, A. & Thornton, G. B. *J. Immun.* **139**, 1223-1231 (1987).
11. Rajewsky, K., Schirmacher, V., Nase, S. & Jerne, N. K. *J. exp. Med.* **129**, 1131-1137 (1969).
12. Russell, S. M. & Liew, F. Y. *Nature* **147**, 280-282 (1979).
13. Lake, P. & Mitchison, N. A. *Cold Spring Harb. Symp. quant. Biol.* **41**, 589-596 (1976).
14. Milich, D. R., McLachlan, A., Moriarty, A. & Thornton, G. B. *J. Immun.* **138**, 4457-4465 (1987).
15. Milich, D. R., McLachlan, A., Chisari, F. V., Kent, S. B. & Thornton, G. B. *J. Immun.* **137**, 315-322 (1986).
16. Milich, D. R. *et al. Science* **228**, 1195-1199 (1985).
17. Milich, D. R., McNamara, M., McLachlan, A., Thornton, G. & Chisari, F. V. *Proc. natn. Acad. Sci. U.S.A.* **82**, 8168-8172 (1985).
18. Milich, D. R. *Vir. Immun.* (in the press).
19. Takai, E. *et al. J. Immun. Meth.* **95**, 23-29 (1986).
20. Itoh, Y. *et al. Proc. natn. Acad. Sci. U.S.A.* **83**, 9174-9178 (1986).
21. Delpeyroux, F. *et al. Science* **233**, 472-475 (1986).
22. Milich, D. R., McLachlan, A., Chisari, F., Nakamura, T. & Thornton, G. B. *J. Immun.* **137**, 2703-2710 (1986).
23. Gerin, J., Faust, R. M. & Holland, P. V. *J. Immun.* **115**, 100-108 (1975).

Directional non-cell autonomy and the transmission of polarity information by the *frizzled* gene of *Drosophila*

Charles R. Vinson* & Paul N. Adler

Biology Department, Molecular Biology Institute and Cancer Center, Gilmer Hall, University of Virginia, Charlottesville, Virginia 22901, USA

The function of the *frizzled* (*fz*) locus is required for the development of a parallel array of bristles and hairs on the adult cuticle of *Drosophila melanogaster*^{1,2}. Marked *fz* mitotic clones from five alleles were generated and examined in the wing. Three alleles have a non-cell-autonomous hair polarity phenotype; wild-type cells distal to *fz* clones produce hairs that have an abnormal polarity. In contrast, *fz* clones of the other two *fz* alleles examined do not disrupt the polarity of neighbouring cells. These data suggest that *fz* has two mutably separate functions in establishing hair polarity on the wing. One function involves the transmittance and/or generation of a polarity signal along the proximal-distal axis of the wing. The second function involves the cellular interpretation of a polarity signal.

The bristles and hairs (and other structures) that decorate the cuticle define a tissue polarity, and have been used as an indicator of fundamental cell polarity³⁻⁷. Bristles (four-cell sensory organs) and hairs (cellular extensions that later become sclerotized) usually point distally on appendages and posteriorly on the thorax and abdomen. We wish to understand how a group of cells, such as the epidermis of *Drosophila*, coordinate their cytoskeletal outgrowths⁸ to produce a parallel array.

Several recessive viable loci in *Drosophila*, termed tissue polarity loci¹, are necessary for the development of normal bristle and hair polarity. Mutations in each locus result in a complex, unique and reproducible polarity pattern. The *fz* gene is unique among the tissue polarity loci in being essential for polarity patterns throughout the adult cuticle. In addition to altering the polarity of bristles and hairs, mutations in all these loci (to varying degrees) cause some cells to elaborate more than one hair. This multiple-hair cell phenotype indicates that the tissue polarity loci do not act later in development during sclerotization, but act instead before, or during, the elaboration of hairs, at a time when a cell expresses its polarity by the elaboration of a hair.

* Present address: Carnegie Institution of Washington, 115 West University Parkway, Baltimore, Maryland 21210, USA

Received 14 May; accepted 13 August 1987.

1. Milich, D. R. & McLachlan, A. *Science* **234**, 1398-1401 (1986).
2. Eddleston, A. L. W. F. & Williams, R. *Lancet* **ii**, 1543-1545 (1974).
3. Mondelli, M. *et al. J. Immun.* **129**, 2773-2778 (1982).
4. Gerety, R. J., Tabor, E., Purcell, R. H. & Tyeryar, F. J. *J. infect. Dis.* **140**, 642-648 (1979).

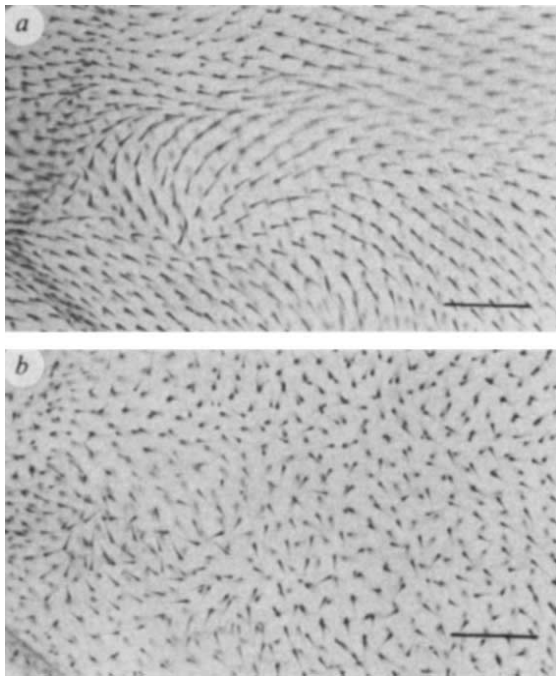


Fig. 1 Hair polarity patterns of *fz* wings. *a*, A weak allele, *fz*^{R53}. *b*, A strong allele, *fz*^{K21}/*fz*^{D21} (inversion/deletion heteroallelic heterozygote). Scale bar, 50 μ m.

A good model system to investigate the generation of polarity patterns is the wing of *Drosophila*. It consists of a dorsal and a ventral sheet of epithelial cells (containing ~26,000 cells per surface) that produce parallel hairs⁹. Weak *fz* alleles cause a swirling of hairs (Fig. 1*a*) in some regions of the wing but have only slight effects in other regions. Strong alleles result in a nearly random orientation of hairs in much of the central region of the wing (Fig. 1*b*), and in noticeable, but less dramatic, alterations in peripheral regions of the wing².

Gubb and Garcia-Bellido¹ generated marked *fz*¹ (*fz*¹ *trc*) mitotic clones in a *Minute* genetic background. They found that large (>100 cells) *fz* clones disrupted the polarity of neighbouring wild-type cells, except when the clones were in the anterior-most cell of the wing. Clones consisting of <50 cells did not affect neighbouring cells. The non-cell-autonomous phenotype seen in large clones is interesting as most loci examined (such as the homeotic loci¹⁰) have a cell-autonomous phenotype¹¹.

We have extended this analysis by generating unmarked and marked *fz* mitotic clones¹² in a wild-type background. Figure 2*a* presents an unmarked *fz*¹ clone. The general location of the *fz* mitotic clone is evident by the appearance of a local disruption of hair polarity (often a swirling type pattern). To mark the boundaries of *fz* mitotic clones, *fz* clones were marked with the gratuitous cell marker *tricorn* (*trc*)¹³. The cellular phenotype of *trc* is a rosette of three or more short hairs, instead of the single long hair seen on the wild-type wing.

Marked *fz* clones were examined for five *fz* alleles (*fz*¹ *trc*, *fz*^{R54} *trc*, *fz*^{KD4a} *trc*, *fz*^{N21} *trc* and *fz*^{F31} *trc*). Mutations in *fz* can be placed into two classes, depending on whether or not they cause a rough eye². Three of the five alleles tested (*fz*¹, *fz*^{R54} and *fz*^{KD4a}) fall into the class that has a rough eye: *fz*¹ is a spontaneous allele used by Gubb and Garcia-Bellido, *fz*^{R54} is our strongest ethylmethane sulphonate (EMS)-induced allele, and *fz*^{KD4a} is a cytologically normal small deletion (derived from hybrid dysgenesis) that removes one of the *fz* exons^{14,15}, and seems to be a phenotypically null mutation. Two of the alleles tested (*fz*^{N21} and *fz*^{F31}) are from the normal-eye class, but have thoracic and wing phenotypes that are more severe than *fz*¹. They were both induced by EMS.

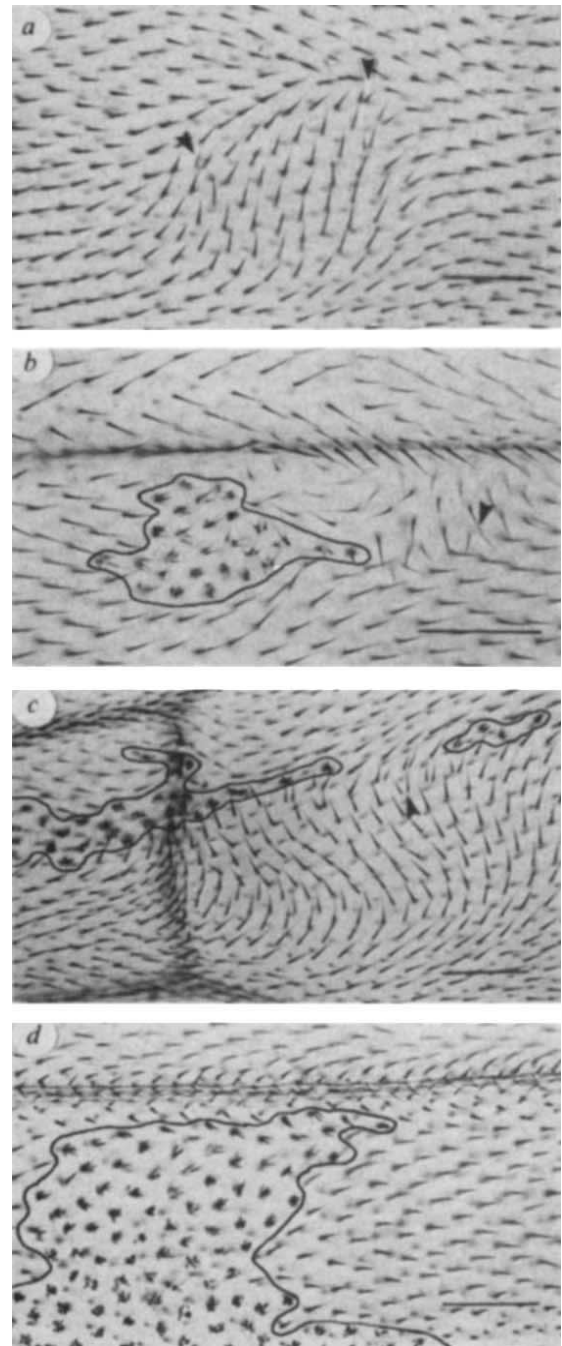


Fig. 2 Appearance of *fz* mitotic clones. *a*, *fz*¹ mitotic clone (induced 2–3 days after egg laying (AEL)). *b*, *fz*¹ *trc* mitotic clone (2–3 days AEL). *c*, *fz*^{R54} *trc* mitotic clone (2–3 days AEL). *d*, *fz*^{N21} *trc* mitotic clone (2–3 days AEL). The *fz* *trc* mitotic clones have been outlined. Arrows in *b* and *c*, presumed wild-type cells that have elaborated two hairs. The *trc* clones did not produce double-hair cells distinct from the clone such as was observed with some *fz* *trc* clones. This suggests that wild-type cells elaborate two hairs as an indirect occurrence of the general polarity disruption. Mitotic clones were induced by irradiation with 1,500 rad (⁶⁰Co source) at the indicated times after egg laying. Eclosed flies were scored for the proper genotype, dehydrated in ethanol, and the wings mounted in Euparal and examined using bright-field optics. Pictures were focused on only one wing surface but remnants of the hairs on the opposite surface are still visible in the micrographs. Scale bar, 50 μ m.

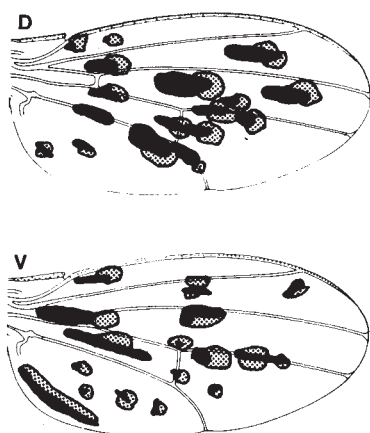


Fig. 3 A composite of the location of $fz^1 trc$ mitotic clones on the dorsal (D) and ventral (V) surface of the *Drosophila* wing. Of 57 clones observed, 34 non-overlapping clones are present on this composite. The black area represents the extent of the $fz trc$ mitotic clone, as marked by trc . The stippled area represents the area of wild-type hair polarity disruption. Mitotic clones induced at 1–2 days AEL, 2–3 days AEL, and 3–4 days AEL are included in the composite. The two putative $fz trc$ mitotic clones (one is present in the composite) that did not disrupt hair polarity could be the result of a double recombination event as trc maps proximally to fz . Presumed fz clones were also observed.

Marked mitotic clones of the three fz alleles of the rough-eye class altered the hair polarity of surrounding wild-type cells (Table 1). In contrast, clones of the two alleles of the normal-eye class did not alter the hair polarity of neighbouring cells. These results suggest that fz has two mutably separate functions in determining hair polarity on the wing.

A careful examination of the wings containing marked fz clones (Fig. 2b and c) indicates a directionality to the non-cell autonomy of rough eye alleles. The polarity of wild-type hairs distal to the mitotic clone is disrupted and they tend to turn in towards the clone. Sometimes the hairs are reoriented 180° and point proximally, not distally. Occasionally, the polarity of wild-type hairs is also disrupted anterior and posterior to the clone. But disruption of the polarity of wild-type hairs is never seen proximal to $fz trc$ mitotic clones, nor on the opposite surface of the wing. The directionality of the non-cell autonomy suggests that fz gene function is required for the transmittance of polarity information along the proximal-distal axis of the wing.

Panels 2b and 2c show a $fz^1 trc$ and a $fz^{R54} trc$ clone respectively. The degree of disruption distal to marked fz clones appeared to be a consequence of two factors, namely the size and shape of clones. Larger clones tended to produce larger regions of disruption. Wide (anterior-posterior axis) or oddly shaped clones produced greater distal disruption than similarly sized, long, thin clones. Distal disruption occurred for marked clones containing as few as ten cells. This result differs from a previous report¹ which could have been due to the use of a *Minute* genetic background, so the small clones observed would

have been induced later in development than those reported here. Panel 2d is a $fz^{N21} trc$ clone. There is no polarity disruption of neighbouring wild-type cells. Such clones are indistinguishable from control trc clones¹⁶.

The location of $fz^1 trc$ mitotic clones and the region of wild-type hair disruption were mapped to the wing to determine if non-autonomy occurred throughout the wing. Figure 3 shows a composite of 34 representative $fz^1 trc$ mitotic clones and the region of disrupted wild-type hair polarity for both the dorsal and ventral surfaces of the wing. Non-cell autonomy of clones occurred throughout the wing. These data suggest that the fz locus is functionally important in all regions of the wing for the generation of the polarity pattern. It is interesting that disruption distal to $fz trc$ clones occurred in regions of the wing that are not dramatically disrupted if the entire wing is fz (such as cell E in ref. 1).

The nature of the signal that is transmitted remains unclear, but there are two distinct possibilities: one is that information molecules are moving along the proximal-distal axis of the wing¹⁷ and the second is that information is moving by cell adhesion, or a cytoskeletal or latticing-type mechanism^{18–20}. These possibilities might be distinguished when the structure of the presumed fz protein is known.

We thank Dr C. P. Emerson for comments on the manuscript. This work was supported by grants from the NIH and NSF. C.V. was supported by an NIH predoctoral training grant.

Received 2 February; accepted 25 August 1987.

- Gubb, D. & Garcia-Bellido, A. *J. Embryol. exp. Morph.* **68**, 37–57 (1982).
- Adler, P., Charlton, J. & Vinson, C. *Dev. Genet.* (in the press).
- Piepho, H. *Biol. Zbl.* **74**, 467–474 (1955).
- Lawrence, P. A. *J. exp. Biol.* **44**, 607–620 (1966).
- Strump, H. *J. exp. Biol.* **49**, 49–59 (1968).
- Nusslein-Volhard, C. & Weischaus, E. *Nature* **287**, 795–801 (1980).
- Weischaus, E. & Riggleman, R. *Cell* **49**, 177–184 (1987).
- Mitchel, H. K., Roach, J. & Peterson, N. S. *Dev. Biol.* **95**, 387–398 (1983).
- Garcia-Bellido, A. & Merriam, J. R. *Dev. Biol.* **24**, 61–87 (1971).
- Morata, G., Sanchez-Herrero, E. & Casanova, J. in *Single Cell Marking and Cell Lineage in Animal Development* (eds Gardner, R. L. & Lawrence, P. A.) 139–144 (Cambridge University Press, 1986).
- Stern, C. in *Genetic Mosaics and Other Essays* (Harvard University Press, 1968).
- Becker, H. J. *Z. Vererb. Lehre* **88**, 333–373 (1957).
- Ferrus, A. thesis, Univ. Autonoma de Madrid (1976).
- Vinson, C. thesis, Univ. Virginia (1987).
- Klein, L. thesis, Univ. Virginia (1987).
- Vinson, C. & Adler, P. *Drosophila Information Service* **66** (in the press).
- Lawrence, P. A. in *Developmental Systems: Insects 2* (eds Counce, S. J. & Waddington, C. H.) 157–206 (Academic, London, 1973).
- Nardi, J. B. & Kafatos, F. C. *J. Embryol. exp. Morph.* **36**, 489–512 (1976).
- Tucker, J. B. *J. Embryol. exp. Morph.* **65**, 1–25 (1981).
- Bissell, M. J., Hall, G. & Parry, G. *J. theor. Biol.* **99**, 31–86 (1982).

The gene for θ -globin is transcribed in human fetal erythroid tissues

Shui-on Leung, Nick J. Proudfoot & Emma Whitelaw

Sir William Dunn School of Pathology, University of Oxford, South Parks Road, Oxford OX1 3RE, UK

Table 1 Percentage $fz trc$ mitotic clones that cause disruption of wild-type hair polarity

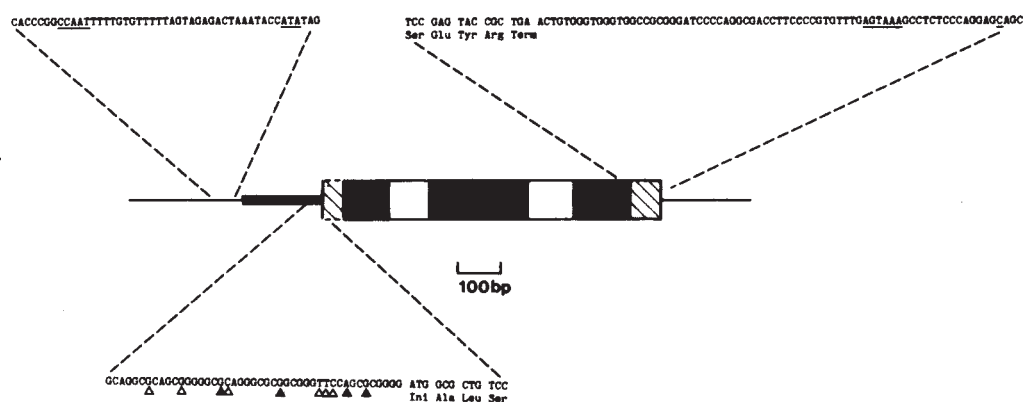
Genotype	Number of $fz trc$ clones identified*	Clones with distal disruption (%)
$fz^1 trc$	57	96
$fz^{KD4a} trc$	33	91
$fz^{R54} trc$	39	77
$fz^{F31} trc$	20	10†
$fz^{N21} trc$	15	6†

* Only clones consisting of more than 15 cells were included.

† The degree of non-autonomy was slight.

A new gene like the α -globin gene has been identified in higher primates at the 3' end of the α -globin gene cluster^{1,2}. There is some controversy as to whether this gene, θ , is a functional globin gene or a non-functional pseudogene^{3–5}. The high degree of sequence conservation displayed by θ between primates^{1,5} and various mammals, such as horse⁶ and rabbit⁷, suggests that this gene is functional in some species. Furthermore, θ encodes a 141-amino-acid polypeptide in sequence similar to α -globin and appears to possess functional RNA-processing signals. But the promoter region of θ is unlike the other globin genes because its CCAAT and ATA box sequences are displaced from the coding sequence by the insertion of a 200-base-pair GC-rich sequence. We demonstrate here the presence of θ -globin messenger RNA in

Fig. 1 The human θ -globin gene, including the 5' and 3' flanking sequences (straight lines), a GC-rich region (solid bar), 5' and 3' non-coding regions and introns (open boxes) and exons (solid boxes). Nucleotide sequences of three regions are shown in the inserts. The first region contains the CCAAT and ATA boxes (underlined), generally found in the promoter regions of globin genes. The second region contains heterogeneous mRNA start sites. These are classified by relative abundance in the mRNA population of K562 cells as either strong ($\blacktriangle$) or weak ($\triangle$). The third region contains the polyadenylation signal, AGTAAA (underlined). The nucleotide to which poly(A) is added is underlined. Intron/exon boundaries are based on those of the baboon and orang-utan¹. DNA sequence analysis was by standard procedures¹⁸.



human fetal erythroid tissue, but not in adult erythroid or other non-erythroid tissues. Furthermore, θ -globin mRNA is detectable in significant amounts in a human erythroleukaemic cell line. These results predict that θ -globin protein will be found in the early stages of human fetal development. Surprisingly, the promoter sequence of θ -globin does not correspond to the CCAAT and ATA box sequences of the gene but rather lies within the adjacent GC-rich sequence, resulting in a heterogeneous series of mRNA 5' ends 50–10 base pairs to 5' of the initiation codon. This type of promoter is reminiscent of that found in housekeeping genes such as adenine deaminase⁹ and hypoxanthine-guanine phosphoribosyl-transferase^{9,10}.

The human θ -globin gene has recently been isolated². Using our knowledge of the sequence of primate θ genes^{1,5}, we identified restriction sites in the human θ gene and then sequenced its 5' and 3'-terminal regions. Comparison of these sequences with those of other primate θ genes^{1,5} shows a high degree of homology: it can be seen from Fig. 1 that the human θ gene contains the CCAAT and ATA promoter sequences at the same location and with the same spatial arrangement, followed by the highly conserved GC-rich sequence, and ending in the initiation codon ATG. The termination codon TGA lies at the expected position and the poly(A) signal is AGTAAA, like the θ gene of the orang-utan and baboon^{1,5}.

The human θ gene, together with 1.3 kilobases (kb) of 5' and 0.4 kb of 3'-flanking sequence, was purified from a cosmid clone (α 3' Bg) (ref. 2) and subcloned into the transient expression vector pSVed, which contains the SV40 origin of replication and enhancer sequences¹¹. θ pSVed was then transfected into HeLa cells, together with a control plasmid which contained the rabbit β -globin and SV40 early genes (R β pBSV). This latter gene encodes the SV40 T antigen required for the replication of both plasmids. Figure 2 compares the expression of θ pSVed with α pSVed, which produces large amounts of α -globin mRNA when transfected into HeLa cells¹². As shown in Fig. 2a, S1 nuclease analysis of cytoplasmic RNA purified from HeLa cells transfected with α - or θ pSVed using 3' α - or 3' θ -DNA probes (Fig. 2c), showed the presence of α -globin mRNA but not θ -globin mRNA. The fact that both transfections gave rabbit β -globin mRNA signals (Fig. 2b) argues that the lack of θ -globin mRNA reflects the inactivity of this gene in HeLa cells. To determine whether the 5' or 3' end of the θ gene was inactive, we constructed α/θ and θ/α gene hybrids fused together at their homologous *Bst*EII restriction enzyme sites at the beginning of the third exon (Fig. 2c). Cytoplasmic RNA from HeLa cells transfected with α/θ -pSVed showed significant levels of α/θ -globin mRNA using the same 3'- θ *Bst*EII probe (Fig. 2a). The 3' end of the α/θ mRNA mapped to about 15 nucleotides 3' of the AGTAAA sequence (see underlined nucleotide in Fig.

1). This indicates that the variant poly(A) site signal, AGTAAA, of θ -globin is functional, which has been shown in other genes^{13,14}. On the other hand, cells transfected with θ/α pSVed gave no detectable θ/α -globin mRNA (Fig. 2a). Two explanations are possible for these results: either the θ promoter is non-functional, defining θ as a pseudogene, or it is tissue-specific and is possibly active only in erythroid cells, like the human ζ -globin gene^{11,15}.

We next investigated θ -globin expression in the human erythroleukaemic cell line K562, which expresses large amounts of ζ - and α -globin mRNA (ref. 16). As seen in Fig. 3a, S1 nuclease analysis with the same 3'- θ -*Bst*EII probe showed evidence for the synthesis of endogenous θ -globin mRNA. This result indicates that the θ gene, at least in man, is a functional gene rather than a pseudogene. We estimate that the amount of θ -globin mRNA produced in K562 cells is about 20% of α -globin mRNA. This value was established by comparing α - and θ -globin mRNA 3' ends using probes of the same specific activity (data not shown). We initially assumed that the CCAAT and ATA boxes of the θ gene would define its promoter region, with the Cap site of the mRNA 30 base pairs (bp) 3' to the ATA box which is 190 bp upstream of the initiation codon. This would give θ -globin mRNA an unprecedented 5' non-coding sequence of 200 GC-rich nucleotides. No detectable 5' ends of θ -globin mRNA were found in this region of the gene (data not shown). Instead, multiple 5' ends of θ -globin mRNA were identified when the GC-rich region close to the initiation codon was used as an S1 nuclease probe with K562 RNA (see Fig. 3b). By purifying the polyadenylated mRNA fraction of K562, these multiple 5' end signals could be identified and precisely mapped by using the DNA sequence analysis fractionated on the same polyacrylamide gel. The full-length probe band in the KA⁺ and K lanes of Fig. 3b reflects the double-stranded nature of this probe, rather than the presence of upstream transcripts initiating 5' to the *Sma*I site, because very little full-length probe signal was detected using a single-stranded 5' θ probe (data not shown). We have confirmed the presence of multiple θ -globin mRNA 5' ends in the -5 to -30 region of the θ gene by using the alternative primer extension analysis (data not shown). The precise positions of these multiple start sites are indicated in Fig. 1. The position and heterogeneity of the 5' termini of θ -globin mRNA makes it unlikely that the θ gene uses the CCAAT and ATA box sequences as promoter because the ATA box directs a unique start site for transcription about 30 bp downstream. We suggest that the promoter of the θ gene lies within the GC-rich region and, like other GC-rich non-TATA box promoters, directs heterogeneous initiation of transcription. The fact that the θ gene possesses a promoter reminiscent of the GC-rich promoters associated with housekeeping genes⁸⁻¹⁰

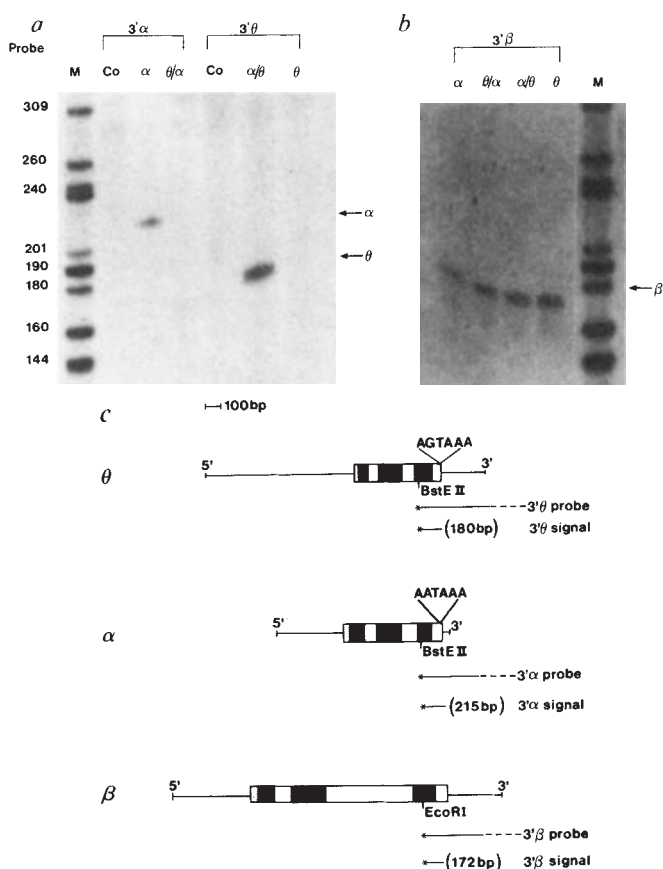


Fig. 2 θ -Globin gene expression in transfected HeLa cells. *a*, S_1 nuclease analysis of cytoplasmic RNA from HeLa cells transfected with plasmids containing the θ -globin gene (θ pSVed), the α_1 -globin gene (α_1 pSVed), the α/θ hybrid (α/θ pSVed), or the θ/α hybrid (θ/α pSVed). θ pSVed consists of a 2.5-kb *EcoRI*/*HindIII* fragment containing human θ -globin gene with 1.3 kb of 5'-flanking sequence and 0.4 kb of 3'-flanking sequence cloned into the *ScaI* site of the expression vector pSVed (ref. 11). α_1 pSVed consists of a 1.6-kb *PstI* fragment containing the human α_1 -globin gene cloned into the *PstI* site of the expression vector pSVed (ref. 12). The third exon of both the θ -globin gene and the α_1 -globin gene contains a *BstEII* site. By cutting both genes with *BstEII* the 5' region of the α_1 -globin gene was ligated to the 3' region of the θ -globin gene (α/θ pSVed) and the 5' region of the θ -globin gene was ligated to the 3' region of the α_1 -globin gene (θ/α pSVed). HeLa cells were transfected with plasmid DNA using calcium phosphate coprecipitation and cytoplasmic RNA was prepared 48 h later¹¹. Cytoplasmic RNA was hybridized to either a double-stranded 3'- θ -globin probe end-labelled with Klenow DNA polymerase at the *BstEII* site or to a double-stranded 3'- α_1 -globin probe end-labelled with Klenow DNA polymerase at the *BstEII* site, and digested with *S1* nuclease¹¹. *b*, HeLa cells were cotransfected with a plasmid R β SV-pBR328 containing both the rabbit β -globin gene and the SV40 large T antigen which allows all pSVed plasmids (containing SV40 ORI) to replicate. *S1* nuclease analysis used RNA hybridized to a double-stranded 3'- β -globin probe end-labelled with Klenow DNA polymerase at the *EcoRI* site. M denotes size markers. *c*, Diagram showing the different gene fragments cloned into the expression vector. The probe used in the *S1* nuclease analysis and the size of the protected fragment after *S1* nuclease treatment is shown below each diagram. * Denotes 32 P-end-label.

suggests a novel form of gene regulation for the gene θ -globin.

We have carried out S_1 nuclease analysis on RNA from a variety of human fetal and adult erythroid and also non-erythroid tissues using the 3'- θ -, α - and ζ -*BstEII* probes (Fig. 4). We find detectable levels of globin mRNA in both fetal yolk sac and fetal liver (Fig. 4a, panels II and III), whereas no

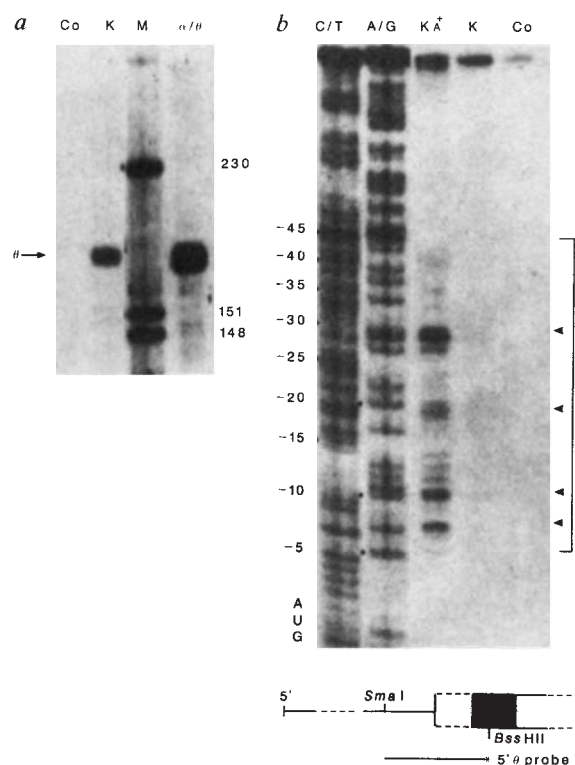


Fig. 3 θ Globin mRNA in human erythroleukaemic cells (K562). *a*, Identification of the 3' end of θ -globin mRNA from K562 cells: Cytoplasmic RNA from K562 cells (K), from HeLa cells transfected with α/θ pSVed as a positive control (α/θ), and from tRNA as a negative control (Co) was hybridized to the *BstEII* end-labelled probe (see Fig. 2c). *S1* nuclease digestion products were run on 4% urea-polyacrylamide gel. M denotes size markers. *b*, Identification of the 5' termini of θ -globin mRNA in K562 cells: Cytoplasmic RNA from K562 cells (10 μ g), poly(A)-selected mRNA (K^+) (5 μ g) or tRNA was hybridized to a kinase-labelled, double-stranded *SmaI*-*Bss*HII fragment from the 5' end of the θ -globin gene (5'- θ -probe), as shown under the figure. *S1* nuclease digestion products were separated on a urea-polyacrylamide gel. Cytoplasmic RNA from K562 cells was poly(A)-selected using an oligo(dT) column¹⁹. Filled-in arrows show the strong start sites of the mRNA. The region over which the multiple start sites are detectable is indicated by a vertical bracket, and their precise positions shown by the sequence ladder¹⁸ of the same 5'- θ -probe run alongside the *S1* digest. The C/T (hydrazine) and A/G (formic acid) degradations are shown. AUG represents the initiation codon position on the sequence ladder. Positions -5 to -45 are those of nucleotides before AUG. * Denotes 32 P-end label.

appreciable θ -globin mRNA was found in either adult erythroid or in non-erythroid tissues at any developmental stage tested (Fig. 4a, panel I). This pattern of θ -globin mRNA expression mimics that of ζ -globin mRNA, as shown in Fig. 4b. It is unlikely that negative results are due to the testing of insufficient RNA because the fetal and adult erythroid tissues gave a comparable signal for α -globin mRNA, the strongest being in bone marrow (Fig. 4b). These results indicate that θ is an early erythroid gene, possibly expressed at a similar developmental stage to the embryonic ζ gene. But although the θ gene is transcriptionally active in human yolk sac, fetal liver and the human erythroleukaemic cell line, no θ -globin polypeptide has so far been reported. It is improbable that a polyadenylated and relatively abundant mRNA would be translationally silent. Indeed, a recent analysis of early embryonic and fetal globins in man¹⁷ showed the presence of additional unassigned polypeptides. One of these unidentified proteins could correspond to θ -globin, but the function of θ -globin in the physiology of human erythroid development remains an enigma.

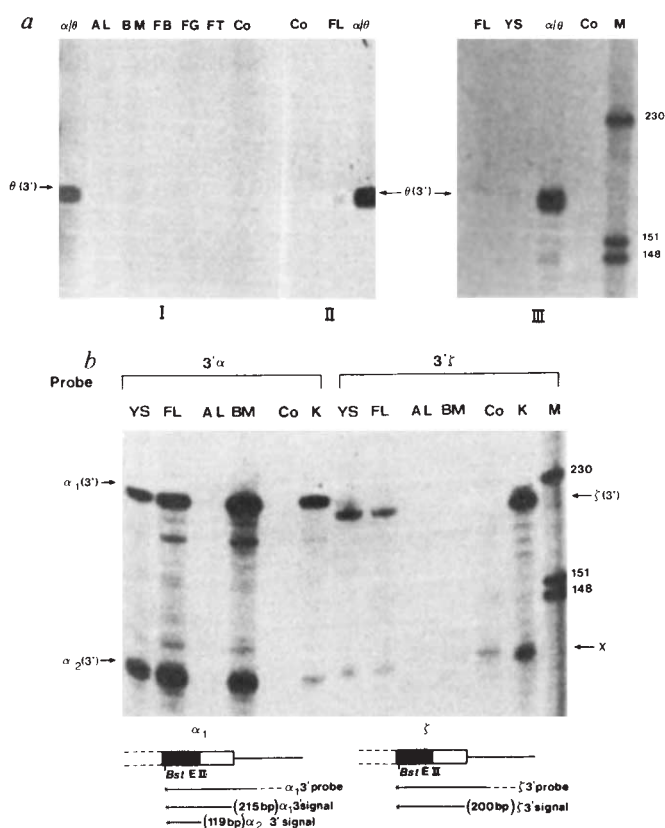


Fig. 4 Analysis of θ -globin mRNA in human erythroid and non-erythroid tissues: adult liver, AL; adult bone marrow, BM; fetal brain, FB; fetal gut, FG; fetal tongue, FT; fetal liver, FL; fetal yolk sac, YS; K562, K; markers, M. *a*, Identification of θ -globin mRNA in fetal erythroid but not in adult erythroid or non-erythroid tissues: S1 nuclease analysis using the *Bst*EII 3'- θ probe of RNA preparations from adult erythroid and non-erythroid tissues: AL, BM, FB, FG and FT give no detectable θ signals (panel I). Cytoplasmic mRNA from HeLa cells transfected with α/θ pSVed acts as a positive control (α/θ). The autoradiograph was heavily exposed to indicate no signals above background, tRNA (Co). Panels II and III show two separate S1 nuclease analyses of mRNA from 6–12 week old fetal liver and fetal yolk sac. A clear θ signal is seen with FL mRNA in Panel II. Panel III shows two faint but positive signals in FL and YS mRNA. *b*, Identification of α - and ζ -globin mRNA in erythroid tissues: S1 nuclease analysis with 3'- α_1 - and 3'- ζ -probes (* denoted 32 P-end label) on mRNA from fetal liver, fetal yolk sac, adult liver, adult bone marrow, tRNA (Co) and K562. The 3'- α_1 -probe gives a signal of 215 bp for α_1 -globin mRNA and a mismatch signal for α_2 -globin mRNA at 119 bp (the position at which α_2 - and α_1 -globin mRNA diverge in sequence in their 3' non-coding regions²⁰). The 3'- ζ -probe gives a signal at 200 bp for ζ -globin mRNA and also a faint artifact signal (denoted by a cross) in the tRNA control lane. α_2 - and α_1 -globin mRNA is detectable in all the erythroid tissues tested (FL, YS, BM, K) and, as predicted, is strongest in bone marrow. ζ -globin mRNA is detectable in fetal liver and yolk sac but not in adult bone marrow.

We thank Dr Doug Higgs for providing the human θ gene, Dr Mark Vickers for the human bone marrow RNA sample, and to Professor Chris Graham for all of the other human erythroid and non-erythroid RNA samples. We also thank Drs John Clegg and Doug Higgs for encouragement. This work was funded by the Medical Research Council (N.J.P.). S.L. is an ICI-Oxford Scholar.

- Proudfoot, N. J. *Nature* **321**, 730–731 (1986).
- North, G. *Nature* **326**, 641 (1987).
- Shaw, J.-P., Marks, J. & Shen, C.-K. *J. Nature* **326**, 717–720 (1987).
- Clegg, J. B., Goodbourn, S. E. Y. & Braend, M. *Nucleic Acids Res.* **12**, 7847–7858 (1984).
- Cheng, J.-K., Raid, L. & Hardison, R. C. *J. Biol. Chem.* **261**, 839–848 (1986).
- Valerio, D. *et al.* *EMBO J.* **4**, 434–443 (1985).
- Melton, D. W., McEwan, C., McKie, A. B. & Reid, A. M. *Cell* **44**, 319–328 (1986).
- Dynan, W. S. *Trends Genet.* **2**, 194–197 (1986).
- Proudfoot, N. J., Rutherford, T. R. & Partington, G. A. *EMBO J.* **3**, 1533–1540 (1984).
- Higgs, D. R. *et al.* *Nature* **306**, 398–400 (1983).
- Capon, D. J., Chen, E. Y., Levinson, A. D., Seeburg, P. A. & Goeddel, D. V. *Nature* **302**, 33–37 (1983).
- Donehower, L., Huang, A. L. & Hager, G. L. *J. Virol.* **37**, 226–238 (1981).
- Norman, C., Raymond, V. & Proudfoot, N. J. *Hemoglobin Switching V, Proc. 5th Conference on Hemoglobin Switching* (Liss, New York, in the press).
- Rutherford, T. R., Clegg, J. B. & Weatherall, D. J. *Nature* **280**, 164–165 (1979).
- Peschle, C. *et al.* *Nature* **313**, 235–238 (1985).
- Maxam, A. M. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* **74**, 560–564 (1977).
- Aviv, H. & Leder, P. *Proc. natn. Acad. Sci. U.S.A.* **69**, 1408–1412 (1972).
- Michelson, A. M. & Orkin, S. H. *Cell* **22**, 371–377 (1980).

Germline mosaicism and Duchenne muscular dystrophy mutations

E. Bakker*, Ch. Van Broeckhoven†, E. J. Bonten*, M. J. van de Vooren‡, H. Veenema*, W. Van Hul†, G. J. B. Van Ommen*, A. Vandenberghe† & P. L. Pearson*

* Department of Human Genetics, State University of Leiden, The Netherlands

† Department of Biochemistry, University of Antwerpen, Belgium

‡ Department of Clinical Genetics, Regional Laboratories 'Zeeland', Middelburg, The Netherlands

Duchenne muscular dystrophy (DMD) is a severe X-linked neuromuscular disease with an incidence of ~1 in 3,500 newborn boys. The DMD locus has a high mutation frequency: one third of the cases is thought to result from a new mutation¹. Linkage studies using probes to detect restriction fragment length polymorphisms² and DNA deletion studies³ have greatly improved DMD carrier detection and prenatal diagnosis^{4,5}. Here we report on two families in which a *pERT87* (*DXS164*) deletion^{3,6} was transmitted to more than one offspring by women who showed no evidence for the mutation in their own somatic (white blood) cells. We also show that the deletion in both siblings in one of the families is identical, indicating that the deletion must have occurred during mitosis in early germline proliferation, leading to a germline mosaicism. This phenomenon may turn out to be a major factor contributing to the induction of DMD mutations, and has important implications for the counselling of DMD families.

The first DMD pedigree described involves a sporadic DMD case, as shown in Fig. 1a. The patient (ind. 10) carries a deletion for the probes p87-1, p87-8 and p87-15 at the *pERT87* (*DXS164*) locus. Of the eight restriction fragment length polymorphisms described previously⁷ and one additional *EcoRV* RFLP found by us⁵, none were found to be informative in the DNA of the patient's mother (ind. 6). Further studies on the patient's DNA revealed that the deletion is delineated by p87-30 and p87-42, which were themselves not deleted. Figure 1a shows the hybridization pattern with p87-1 and C7 after *EcoRV* digestion of DNA from members of this family. On the basis of the relative band intensities it is obvious that, with the exception of the DMD patient (ind. 10), who lacks the p87-1 allele of the mother, all individuals have a normal copy number for alleles of both X-probes. In particular, the intensity of the signal of the P allele given by p87-1 in the mother (ind. 6) of the patient when compared to that from normal family members strongly suggests that the deletion is not present in maternal DNA isolated from white blood cells. On this basis the patient's mother should be considered as a non-carrier (her CK-level was normal). But to avoid depending solely on the result of quantitative

Received 20 July; accepted 7 September 1987.

- Marks, J., Shaw, J.-P. & Shen, C.-K. *J. Nature* **321**, 785–788 (1986).
- Nicholls, R. D., Fischel-Ghodsian, N. & Higgs, D. R. *Cell* **49**, 369–378 (1987).

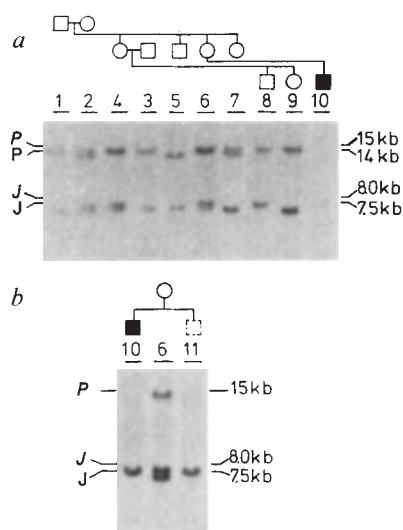


Fig. 1 *a*, Southern blot of *EcoRV*-digested DNA from family MD6. DNA analysis⁵ of an isolated case of DMD shows a p87-1 deletion in the proband (ind. 10). The probes p87-8 and p87-15 were also deleted (data not shown). Combined hybridization of the probes p87-1 and C7 to an *EcoRV* blot, shows the quantitative analysis of the DNA of the patient's mother (ind. 6). If the hybridization intensities of the bands in lanes corresponding to the individuals 2, 4 and 6 are compared it is clear that individuals 4 and 6 have a double intensity for the upper allele of p87-1 (*P*) and are not carrying the deletion. *b*, Prenatal diagnosis in family MD 6 using an *EcoRV* Southern blot hybridized to probes p87-1 and C7. The male fetus (ind. 11) clearly lacks the p87-1 signal which is also absent in the patient's DNA (ind. 10), Fig. 1*a*.

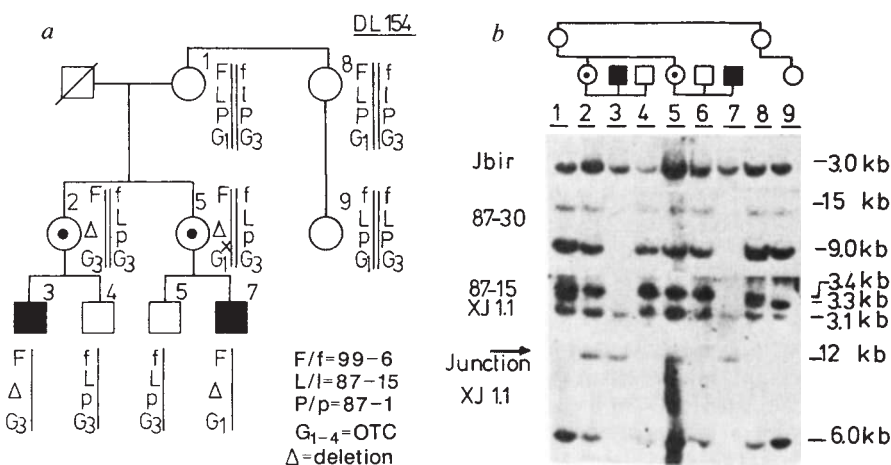
blots for carrier detection, we advised prenatal diagnosis of the next pregnancy as the DMD deletion mutation would be directly detectable. Fetal DNA analysis revealed that the same *pERT87* markers (p87-1, p87-8 and p87-15) were absent in the male fetus (ind. 11, see Fig. 1*b*), and that p87-30 and p87-42 were present. This result was identical to that from the proband and termination of the pregnancy was advised.

A similar situation presented itself as a straightforward familial case of DMD (Fig. 2*a*). On the basis of pedigree data the grandmother (ind. 1) of the two DMD patients (inds 3 and 7) would formally be designated as a carrier. DNA analysis using the RFLP probes, described in our diagnostic procedure⁵, demonstrated a deletion spanning the whole *pERT87* locus in

individuals 3 and 7. In this situation, the grandmother could not possibly carry the *pERT87* deletion in her leucocytes because she was heterozygous for the probe p87-15, located within the patient's deletion. The Southern blot analysis shown in Fig. 2*b* demonstrates an absence of hybridization for alleles of p87-30 and p87-15 in the two patients. Moreover, both patients and their mothers show an identical *PstI* junction fragment of 12 kilobases (kb) with the probe XJ 1.1 (ref. 8) (Fig. 2*b*) that is not present in the DNA of the grandmother. This strongly suggests that both patients have inherited an identical deletion chromosome, although final proof would require cloning and sequence analysis of the junction point. As this deletion, which should span 220 kb on the basis of the markers deleted⁷, is not present in the grandmother's leucocytes (somatic cells), a single deletion-mutation has probably occurred in oögonia, leading to gonadal mosaicism and to more than one affected germline cell.

Several other cases suggesting germline mosaicism for the DMD mutation have been reported: one family was described by Monaco *et al.*⁷, a second by Lanman *et al.*¹⁸, and a third by Kaplan and coworkers (personal communication). A similar case has been described for Becker muscular dystrophy⁹. But in these cases the phenomenon could only be inferred and was not confirmed through death of one of the affected individuals before analysis. These findings have implications for genetic counselling; it is imperative to reconsider the recurrence risk in all those cases which appear to be new mutations. In 102 unrelated families (92 Dutch and 10 Belgian families) analysed during the last two years, 12 involved detectable deletions, two of which were recurrent new 'mutations'. This phenomenon could be present in the remaining families, but cannot yet be demonstrated because the DMD mutation itself remains undetected. The conclusion is that germline mosaicism could be a major contributing factor to multiple cases of DMD within one family. Reviewing our pedigree and/or RFLP data 15 women from the 102 families can be pinpointed who were previously thought to be normal, but must now be considered to have an increased risk of being a DMD carrier. The use of more probes and alternative methods of analysis will permit direct detection of a higher fraction of mutations and hence these observations can be extended to other families at risk. Also, prenatal diagnosis should be offered to all potential carriers of this type, namely mothers of isolated DMD patients who were found not to carry a deletion mutation present in their male offspring. In addition, carrier detection should be performed on the sisters of such males. In the analysis of familial cases with affected males in more than one generation the

Fig. 2 *a*, Pedigree of family DL 154. A familial case of DMD, two cousins (ind. 3 and 7) are affected, their mothers obligate carriers (ind. 2 and 5) and the grandmother (ind. 1) a presumptive obligate carrier. DNA analysis⁵ detects the DMD mutation as a deletion for the entire *pERT87* locus. The grandmother (ind. 1) appears not to be deleted, being heterozygous for p87-15 (L1) and having a double intensity for the p87-1 (PP) hybridization signal (data not shown) (ind. 2 and 5 show that the deletion is of grandmother origin). *b*, Southern blot analysis of the deletion. Hybridization of the Xp212 DMD-specific probes on family DL154 blot show that the deletion is flanked by Jbir on the distal side (top autoradiograph) and XJ1.1 on the proximal side (3.1 kb band). The deleted region overlaps 87-30 (*PvuII* digestion, second autoradiograph) and 87-15 (*TaqI* digestion, 3.4 and 3.3 kb alleles). Both carrier women (inds. 2 and 5) show only one copy of p87-15, relative to the hybridization signal of XJ1.1. XJ1.1 is not deleted in the DNA of both patients. Other probes tested were 87-42, XJ 10.1, XJ 8.2 and XJ 2.3 and were also deleted (data not shown) in >200 kb. The bottom panel shows a *PstI* digest in which the probe XJ 1.1 detects a 12 kb *PstI* fragment in both patients (lanes 3 and 7), instead of the normal 6 kb fragment. Both mothers of the patients (lane 2 and 5) also show this 12 kb deletion junction fragment, whereas the grandmother (lane 1) is homozygous for the normal 6 kb XJ 1.1 band. The weak additional band in lane 8 between 6 and 12 kb is due to incomplete digestion.



DMD-mutation may not be detectable directly, but the transmission of affected X-chromosomes can nearly always be followed by the use of DMD flanking and internal RFLPs.

An alternative to the germline mosaicism explanation advanced here is that two consecutive independent but identical mutations have occurred, which we consider to be far less likely. It is apparent that DMD mutations, as a whole, are widely spread both in location and size and that mutations are not localized to specific positions within the gene¹⁰.

There is no reason to expect that the mitotic events causing germline mosaicism should be limited to the DMD gene locus. Gitschier *et al.*¹¹ recently described a case in which unusual transmission of mutations in the gene for clotting factor VIII is consistent with gonadal mosaicism. Earlier studies by Hartl *et al.* and Murphy *et al.*^{12,13} considered the demonstration of gonadal mosaicism for lethal recessive X-linked disorders and calculated the increased recurrence risks induced by this phenomenon. Gartler and Francke^{14,15} have suggested that the induction of mosaicism by half chromatid lesions could explain a possible excess of carriers in Lesch-Nyhan's disease. Although molecular evidence for such a mechanism in man has not been forthcoming, it must be considered as a possible cause of germ-

line mosaicism. Alternative mechanisms proposed recently^{16,17} involve unequal recombination and gene conversion that can give rise to both somatic and germline mosaicism.

We thank Drs R. A. Worton, L. M. Kunkel, W. A. Fenton and J. L. Mandel for supplying several of the DNA probes used. This work was financially supported by the Dutch Prevention Fund and by the Muscular Dystrophy Group of Great Britain.

Received 22 May; accepted 21 August 1987.

1. Moser, H. *Hum. Genet.* **66**, 17-40 (1984).
2. Davies, K. E. *et al. Nucleic Acids Res.* **11**, 2303-2312 (1983).
3. Kunkel, L. M. *et al. Proc. natn. Acad. Sci. U.S.A.* **82**, 4778-4782 (1985).
4. Bakker, E. *et al. Lancet* **i**, 655-658 (1985).
5. Bakker, E. *et al. J. med. Genet.* **23**, 573-580 (1986).
6. Kunkel, L. M. *et al. Nature* **322**, 73-75 (1986).
7. Monaco, A. P. *et al. Hum. Genet.* **75**, 221-227 (1987).
8. Ray, P. N. *et al. Nature* **318**, 672-675 (1985).
9. Romeo, G. *et al. Am. J. hum. Genet.* **39**, 291 (1986).
10. Koenig, M. *et al. Cell* **50**, 509-517 (1987).
11. Gitschier, A. J. *Am. J. hum. Genet.* **39**, 593 (1986).
12. Hartl, D. L. *Am. J. hum. Genet.* **23**, 124-134 (1971).
13. Murphy, E. A. *et al. Am. J. hum. Genet.* **26**, 207-222 (1974).
14. Gartler, S. M. & Francke, U. *Am. J. hum. Genet.* **27**, 218-223 (1975).
15. Francke, U. *et al. Am. J. hum. Genet.* **28**, 123-137 (1976).
16. Thompson, C. B. & Neiman, P. E. *Cell* **48**, 369-378 (1987).
17. Herrmann, B. *et al. Cell* **44**, 469-476 (1986).
18. Lenman, J. T. *et al. Am. J. hum. Genet.* **41**, 138-144 (1987).

A partial deletion of the muscular dystrophy gene transmitted twice by an unaffected male

Basil T. Darras & Uta Francke

Department of Human Genetics, Yale University School of Medicine, New Haven, Connecticut 06510, USA

A gene of unknown function located in band Xp21 on the short arm of the human X chromosome gives rise to X-linked recessive muscular dystrophy, of either Duchenne or Becker type, when mutated. The gene encodes a large muscle-specific transcript of about 14 kilobases (kb) (refs 1 and 8) and its genomic size extends over more than 1,800 kb (refs 2 and 3). The high mutation rate (about 10^{-4} per generation)⁴ is likely to result from the large target size. Submicroscopic deletions, detectable with one or more of the dozen cloned DNA probes available for regions within the gene, constitute a significant proportion of the mutations^{3,5}. Because no such deletions have been found in normal individuals, it is assumed that intragenic deletions are the molecular basis of the mutations. The origin of deletions can be traced in families. With sufficient data collected, it will soon be possible to answer questions about the relative frequencies of mutations in male and female gametogenesis and about the timing of mutational events in mitotic or meiotic stages of germ cell development. We have studied a four generation family containing males who have Duchenne muscular dystrophy due to deletion of the sequence recognized by intragenic probe J-Bir (ref. 6). The deletion was present in two of five daughters of a woman who herself did not have the deletion. Haplotype analysis on 15 members of this family using nine informative restriction fragment length polymorphism (RFLP) markers indicated that the J-Bir deletion chromosome was transmitted from the unaffected father.

The family came to us for prenatal diagnosis when individual III-2 (see Fig. 1) was pregnant with a male fetus (IV-1). DNA analysis of a chorionic villus sample revealed the J-Bir deletion and a first-trimester termination ensued. The J-Bir deletion was present in two sisters of generation II (II-2 and II-4) who were considered proven carriers because each had a son with classical Duchenne muscular dystrophy (DMD). The deletion was also found in the surviving affected male (III-1). It was not present, however, in his grandmother (I-1) who would have been con-

sidered an obligate carrier by conventional pedigree analysis. Instead, I-1 was heterozygous for the RFLP detected by probe J-Bir.

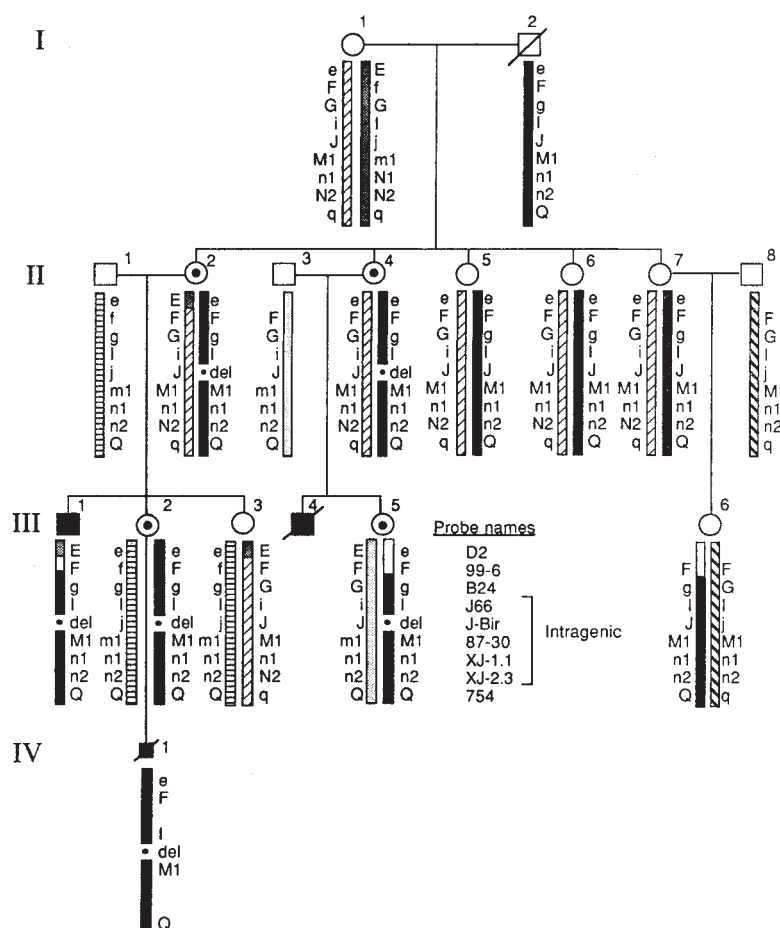
To determine which grandparent had transmitted the X chromosome with the deletion to two daughters, we carried out haplotype analysis with nine informative RFLP markers from the Xp21-p22 region (Table 1). Probes D2, 99-6 and B24 map distal to the DMD locus, 754 is a proximal flanking marker and XJ1.1, XJ2.3, 87-30, J-Bir and J-66 are from within the gene^{2,6,7}. Haplotypes were parsimoniously assigned as shown in Fig. 1. No recombination events were detected between the closest flanking markers B24 and 754. Although individual I-2 was dead, his haplotype could easily be inferred by the results obtained in the females in two subsequent generations. The hypothetical grandpaternal haplotype is indicated in black in Fig. 1. The results clearly demonstrate that this was the chromosome in which the J-Bir deletion arose. When individual I-2 had died in his sixties after a myocardial infarct, he had no history of muscle weakness.

It was surprising that all five daughters in generation II had received the same Xp21 region from their mother. We have confirmed by using high-frequency autosomal RFLPs that the samples studied were from different individuals (data not shown). For genetic counselling we have identified female carriers of the deletion (III-2 and III-5) who run a high risk of producing DMD offspring and have been able to reassure two others (III-3 and III-6) that they are not at risk because they both are heterozygous for the J-Bir RFLP.

The significant observation is the finding that two out of five X chromosomes transmitted by an unaffected male carried the same partial deletion of the DMD gene. Studies with complementary DNA probes covering the deleted region⁸ have confirmed that the deletions in II-2 and II-4 are identical. The restriction fragments missing in affected males are of single-copy intensity in all carrier females. This striking result points to a type of mutational event rarely documented before in the human species. This deletion probably arose at an early mitotic stage in germ cell lineage development, as it is unlikely that the identical products of a late mitotic or single meiotic mutation during spermatogenesis would be detectable in two daughters born seven years apart. Monaco *et al.*⁶ have reported the apparent mitotic origin of a similar deletion mutation in two sisters, but their limited data did not permit a conclusion regarding the parental origin of the deletion.

In the early stages of embryonic development primordial germ cells undergo mitotic divisions and migrate to the undifferenti-

Fig. 1 Pedigree of the family studied. Squares, males; circles, females. Filled symbol, affected individual; open symbol, unaffected; circle with dot, carrier. Diagonal line, deceased. Also shown for each individual are DNA polymorphisms determined with the Xp21-p22 derived probes listed in Table 1. All are simple two-allele RFLPs. Upper case letters indicate the larger restriction fragment and lower case letters the smaller allelic fragment for each. The J-Bir deletion is indicated by the symbol 'del'. The polymorphic sites are represented in their established linear order in telomeric (top) to centromeric (bottom) orientation². They span five chromosomal sub-bands on the mid-portion of the human X chromosome short arm (Table 1). Haplotypes in females were assigned by assuming a minimum number of cross-overs and taking male haplotypes into account. Individual haplotypes are depicted by specific shading or hatching. White areas for D-2 and 99-6 alleles in III-1, III-5 and III-6 indicate uncertainty about their chromosomal origin as the respective mothers were homozygous. Because I-2 is dead, his haplotype had to be inferred. All other results were obtained by Southern blot analysis of DNA extracted from whole blood, lymphoblastoid cell lines or chorionic villi (IV-1) using previously described methods²². It is apparent that the haplotype indicated in black exists in two versions: with (II-2, II-4, III-1, III-2, III-5 and IV-1) and without (II-5, II-6, II-7 and III-6) the J-Bir deletion (-del), and that this haplotype came from male (I-2).



ated gonad. Deletions occurring at that time are likely to produce germinal mosaicism in both sexes. The determination of male-female origin of such mosaicism has profound implications for the molecular mechanisms to be considered.

The model of non-homologous crossing-over between two X chromosomes that are not lined up accurately—possibly facilitated by interspersed repetitive sequences—could apply to mitotic as well as meiotic errors, but is obviously limited to female gametogenesis. In males, this mechanism is important, but only in the pairing region of the X and Y chromosomes and has been invoked to explain the (meiotic) origin of XX males and XY females⁹. The region of the DMD locus in Xp21 is not known to have a homologous counterpart on the Y chromosome. Thus, interstitial deletions removing several hundreds of kilobases from the DMD gene are not likely to result from interactions between the X and Y chromosomes.

Unequal sister chromatid exchange during DNA replication

(S-phase) or G₂-phase of a mitotic cycle remains a possibility. This mechanism would result in partial duplications as well as deletions. Partial duplication of the DMD gene has been observed in affected boys¹⁰, but grandpaternal origin of such a mutation has not been reported.

The splicing out of the deleted segment during mitotic stages of early male germ cell development could be related to the organization of chromatin loops of different size that are held together by scaffolding proteins¹¹. Reactivation of the X chromosome in early primordial germ cells could well be associated with alterations in chromatin configuration. The exact size of this deletion and the nature of the sequences around the deletion breakpoints will be determined using pulsed-field gel electrophoretic analysis and by cloning and sequencing the deletion endpoints. Scaffolding-associated regions of chromatin loops have been defined in *Drosophila* and may contain topoisomerase II consensus sequences¹².

Table 1 Informative DNA polymorphisms on the short arm of the X chromosome in region Xp21 → p22

Sub-band location*	Locus symbol	Probe name	Restriction enzyme	Allele symbols	Fragment sizes (kb)
Xp22.2 → Xp22.1	<i>DXS43</i>	D2	<i>PvuII</i>	E, e	6.6, 6.0
Xp22.2 → Xp22.1	<i>DXS41</i>	99-6	<i>PstI</i>	F, f	23, 13
Xp21.3	<i>DXS67</i>	B24	<i>BstNI</i>	G, g	1.8, 1.7
Xp21.3	<i>DMD</i>	J-66	<i>PstI</i>	I, i	1.5, 1.3
Xp21.3	<i>DMD</i>	J-Bir	<i>BamHI</i>	J, j	21, 5
Xp21.3	<i>DMD</i>	87-30	<i>BglII</i>	M1, m1	30, 7.7
Xp21.3 → Xp21.2	<i>DMD</i>	XJ-1.1	<i>TaqI</i>	N1, n1	3.8, 3.1
Xp21.3 → Xp21.2	<i>DMD</i>	XJ-2.3	<i>TaqI</i>	N2, n2	7.8, 6.4
Xp21.1	<i>DXS84</i>	754	<i>PstI</i>	Q, q	12, 9.0

* Linear order: telomere-centromere.

Probes J-66 to XJ-2.3 cover more than two-thirds of the DMD gene.

It remains possible that male I-2 was a mosaic for the deletion not only in his germ line but also in his somatic cells. In that case, he would still be clinically unaffected as are most heterozygous females. Such mosaicism could be caused by a mutational event in the early postzygotic embryo or by transmission of a half-chromatid mutation from his mother. If the deletion had occurred during the last DNA replication of the oocyte that gave rise to individual I-2, a half chromatid mutation could have been transmitted that would have produced mosaicism of germinal and somatic tissues in this individual. The hypothetical mechanism of 'half-chromatid mutation', originally suggested for single base substitutions¹³, has recently been extended to the origin of deletion mosaics¹⁴. Its occurrence in humans has not yet been proven. Deletion mosaicism in somatic tissues can now be detected by finding fragments of aberrant size using flanking probes.

Lastly, I-2 could have been a chimaera, that had arisen by spontaneous aggregation of two male zygotes, one carrying the normal X chromosome, the other one carrying the newly mutant X with the deletion. No support for this hypothesis was found by studying autosomal RFLPs in the females of generation II.

Germinal mosaicism is not limited to humans nor to X-linked recessive mutations, but has been reported in *Drosophila*¹⁵, mice¹⁶ and in human autosomal genes¹⁷. The most convincing examples involve the recurrence of new autosomal dominant mutations in siblings born to unaffected parents for disorders with clear-cut phenotypes and complete penetrance such as achondroplasia¹⁸. Furthermore, Rett syndrome presumably due to an X-linked dominant mutation, which is lethal in males, has occurred in sisters born to unaffected mothers, with germinal mosaicism being the most likely explanation¹⁹. Germinal mosaicism for chromosomal translocations has been postulated in families in which more than one offspring had Down syndrome due to the presence of a Robertsonian translocation that was not found in the parents' karyotypes²⁰; and in a parent of two sisters with identical interstitial 16q deletions associated with distinct phenotypic abnormalities²¹.

Recent unpublished observations (for review see ref. 3) have suggested that mitotic origin of DMD deletion mutations may be frequent enough to warrant taking it into consideration when examining and counselling families with muscular dystrophy. A parent in whom germinal mosaicism has been identified has a high risk of producing abnormal offspring¹⁷.

We thank Barbara R. West and Angela Scioscia for obtaining pedigree data and blood samples, Brigitte Foellmer for technical assistance and Louis M. Kunkel, Ron Worton and Gert-Jan van Ommen for DNA probes. This work was supported by an NIH Genetics Center Grant and by the Muscular Dystrophy Association of America.

A mitochondrial target for double-stranded RNA in diseased isolates of the fungus that causes Dutch elm disease

H. J. Rogers*‡, K. W. Buck*§ & C. M. Brasier†

* Department of Pure and Applied Biology, Imperial College of Science and Technology, Prince Consort Road, London SW7 2BB, UK

† Forest Research Station, Alice Holt Lodge, Farnham, Surrey GU10 4LH, UK

The current devastating epidemics of Dutch elm disease in Europe, Southwest Asia and North America are caused by the emergence and spread of a highly pathogenic, aggressive sub-group of the fungus *Ophiostoma* (= *Ceratocystis*) *ulmi* (Buisman) Nannf, which is rapidly replacing the 'old' non-aggressive sub-group, believed to be responsible for the milder disease epidemics earlier in the century¹. There are no effective methods for controlling the spread of the disease nationally, but the discovery of cytoplasmically transmissible factors (d-factors) which cause debilitating diseases in both aggressive and non-aggressive isolates of *O. ulmi* has opened up the possibility of biological control²⁻⁴. One of these, the d²-factor, is associated with specific segments of double-stranded RNA (dsRNA)⁵. We now report that the dsRNA segments in d²-infected isolates copurify with the mitochondria and that mitochondria isolated from such diseased isolates are deficient in cytochrome *aa*₃.

Log1/3-8d² is a diseased (d²-infected) *O. ulmi* isolate which contains ten dsRNA segments, numbered 1 to 10 in order of decreasing size (3.5 kilobase pairs (kb) to 0.33 kb)⁵. Pairs of healthy and d²-infected isolates which were isogenic in their nuclear genomes were produced either by selecting healthy single conidial progeny from Log1/3-8d² or by transferring the d²-factor to healthy isolates by hyphal anastomosis with either donor or recipient carrying a nuclear marker for MBC fungicide tolerance^{5,6}. Their origins and properties are shown in Table 1 and their ratios for cytochromes *aa*₃, *b*, and *c* + *c*₁ in Table 2.

In the three healthy isolates, Con18, H351 and Asc201, all the corresponding cytochrome ratios were similar. In comparing cytochrome ratios of healthy and diseased isolates, only small and inconsistent differences were observed in the *b*/(*c* + *c*₁) ratios. However the ratios of cytochrome *aa*₃ to *b* and (*c* + *c*₁) in the diseased isolates were consistently lower than those in the healthy isolates. For example, the cytochrome *aa*₃/*b* ratios in the first four d²-infected isolates (Table 2) were 3 to 12 times lower than those of the corresponding isogenic healthy isolates. Furthermore the loss of the d²-factor from Log1/3-8d² during conidiogenesis to form Con18, which was accompanied by an increase in the *aa*₃/*b* ratio from 0.26 to 0.79, and the acquisition of the d²-factor by Con18 to form Con18d² (Table 1), which was accompanied by a decrease in *aa*₃/*b* ratio from 0.79 to 0.13, establish the d²-factor (or other cytoplasmic factor lost and gained in the same way) as the cause of the lowered *aa*₃/*b* ratio because Log1/3-8d², Con18 and Con18d² have identical nuclear genomes. Hence lowered levels of the cytochrome *aa*₃ complex, the major terminal oxidase in the respiratory chain in mitochondria of eukaryotes, including fungi, may be a key factor determining the diseased phenotype of d²-infected isolates; note that Con6d², in which cytochrome *aa*₃ was barely detectable, was the most severely debilitated of the isolates tested. An alternative antimycin A and azide-resistant electron transport pathway has been found in *O. ulmi* conidia⁷ and it is possible that respiration in Con6d² proceeds by this alternative pathway.

Received 22 June; accepted 12 August 1987.

1. Monaco, A. P. *et al. Nature* **323**, 646-650 (1986).
2. Monaco, A. P. & Kunkel, L. M. *Trends Genet.* **3**, 33-37 (1987).
3. Robertson, M. *Nature* **327**, 372-373 (1987).
4. Moser, H. *Hum. Genet.* **66**, 17-40 (1984).
5. Kunkel, L. M. *et al. Nature* **322**, 73-77 (1986).
6. Monaco, A. P., Bertelson, C. J., Colletti-Feener, C. & Kunkel, L. M. *Hum. Genet.* **75**, 221-227 (1987).
7. Darras, B. T., Harper, J. F. & Francke, U. *New Engl. J. Med.* **316**, 985-992 (1987).
8. Koenig, M. *et al. Cell* **50**, 509-517 (1987).
9. Page, D. C. *Cold Spring Harb. Symp. quant. Biol.* **51**, 229-235 (1986).
10. Bertelson, C. J. *et al. J. med. Genet.* **23**, 531-537 (1986).
11. Gasser, S. M. & Laemmli, U. K. *Trends Genet.* **3**, 16-22 (1987).
12. Gasser, S. M. & Laemmli, U. K. *Cell* **46**, 521-530 (1986).
13. Gartler, S. M. & Francke, U. *Am. J. hum. Genet.* **27**, 218-223 (1975).
14. Cantu, J. M., Rivas, F., Rivera, H. & Ruiz, C. *Ann. Genet.* **28**, 73 (1985).
15. Hartl, D. L. & Green, M. M. *Genetics* **65**, 449-456 (1970).
16. Dunn, L. C. & Gluecksohn-Waelsch, S. *Genetics* **38**, 261-271 (1953).
17. Hartl, D. L. *Am. J. hum. Genet.* **23**, 124-134 (1971).
18. Frys, J. P., Kleczkowska, A., Verresen, H. & van den Berghe, H. *Clin. Genet.* **24**, 156-162 (1983).
19. Comings, D. E. *Am. J. med. Genet.* **24**, 383-388 (1986).
20. Hall, B. D. *Am. J. med. Genet.* **22**, 417-418 (1985).
21. Hoo, J. J., Lowry, R. B., Lin, C. C. & Haslam, R. H. A. *Clin. Genet.* **27**, 420-426 (1985).
22. de Martinville, B., Blakemore, K., Mahoney, M. J. & Francke, U. *Am. J. hum. Genet.* **36**, 1357-1368 (1984).

‡ Present address: Department of Molecular Genetics, Plant Breeding Institute, Cambridge CB2 2LQ, UK.

§ To whom correspondence should be addressed.

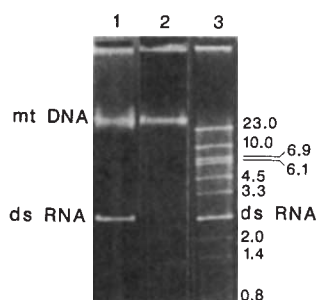


Fig. 1 Double-stranded RNA in a mitochondrial DNA preparation from isolate Con18. Preparation of DNA from purified mitochondria (mtDNA) by centrifugation through a CsCl step gradient separates DNA from single-stranded RNA but an intense band of dsRNA remains (lane 1), which is removed by incubation with ribonuclease (lane 2), but is resistant to digestion with *EcoRI* (lane 3).

Methods. Mitochondrial pellets, prepared from isolate Con18 as in Fig. 2, were resuspended in TNE buffer (50 mM Tris-Cl, 5 mM EDTA, 50 mM NaCl, pH 8.0), containing 2% SDS, incubated at 60 °C for 2 h and then, after addition of nuclease-free pronase (2 mg ml⁻¹), at 37 °C for 2 h. The lysate was extracted several times with a mixture of equal volumes of water-saturated phenol containing 0.1% 8-hydroxyquinoline and chloroform-isoamyl alcohol (24:1, v/v) and then mixed with an equal volume of a CsCl solution in the TNE buffer (density 1.75 g ml⁻¹) and layered on top of the same total volume of the CsCl solution (density 1.75 g ml⁻¹) in a centrifuge tube to generate a step gradient. (Each layer also contained 10 µg ml⁻¹ ethidium bromide.) After centrifugation in a Beckman SW50.1 rotor at 114,000g and 20 °C for 15 h, the nucleic acid band which formed at the interface between the two CsCl concentrations, visualized by ultraviolet fluorescence, was extracted several times with water-saturated butanol to remove ethidium bromide and, after dilution with 3 vols water and precipitation with 2.5 vols ethanol, was resuspended in H₂O. Electrophoresis was in 0.6% agarose gel in 0.09 M Tris, 0.09 M boric acid, 2.5 mM EDTA, pH 8.2, containing 0.5 µg ml⁻¹ ethidium bromide, and nucleic acid bands were detected by ultraviolet fluorescence. Samples were pre-incubated with: lane 2, DNase-free RNase A (100 µg ml⁻¹) in 2 mM Tris-Cl; 0.02 mM EDTA (adjusted to pH 8.2 with acetic acid) at 37 °C for 2 h, conditions shown to digest completely *Penicillium stoloniferum* viral dsRNA (ref. 18); lane 3, *EcoRI* (5 units) in medium salt buffer containing 5 mM spermidine and 100 µg ml⁻¹ nuclease-free bovine serum albumin as described¹⁹. Sizes of restriction fragments (in kb) were estimated relative to *HindIII*-digested λ DNA markers²⁰ (not shown).

Reductions in the levels of cytochrome *aa*₃, sometimes accompanied by changes in the levels of the other cytochromes, are characteristic of cytoplasmically transmissible diseases of other fungi, such as the poky and stopper phenotypes of *Neurospora crassa*, the ragged phenotype of *Aspergillus amstelodami* and senescence in *Podospira anserina*, and are accompanied by deletions in mitochondrial (mt) DNA (for review see ref. 8). Digestion of mtDNA from isolate Log1/3-8d² and diseased and healthy single conidial derivatives (Con6d² and Con18 respectively) with *EcoRI* gave the same pattern of nine fragments (shown for Con18 in Fig. 1), eliminating deletions in mtDNA (except possibly small ones) as the cause of the d²-phenotype. However, preparations of mtDNA contained other major nucleic acid segments, resistant to *EcoRI* but susceptible to ribonuclease digestion, subsequently identified as the dsRNA segments characteristic of each isolate (Table 1). The major band of dsRNA detected in mtDNA preparations from Con18 (Fig. 1) was identified, after separation of the DNA and dsRNA by cellulose chromatography⁹, as dsRNA segment 6 by polyacrylamide electrophoresis⁵, which also revealed small amounts of dsRNA segment 2.

To determine its subcellular distribution, dsRNA in the various subcellular fractions generated during isolation and purification of mitochondria from isolate Log1/3-8d² was ana-

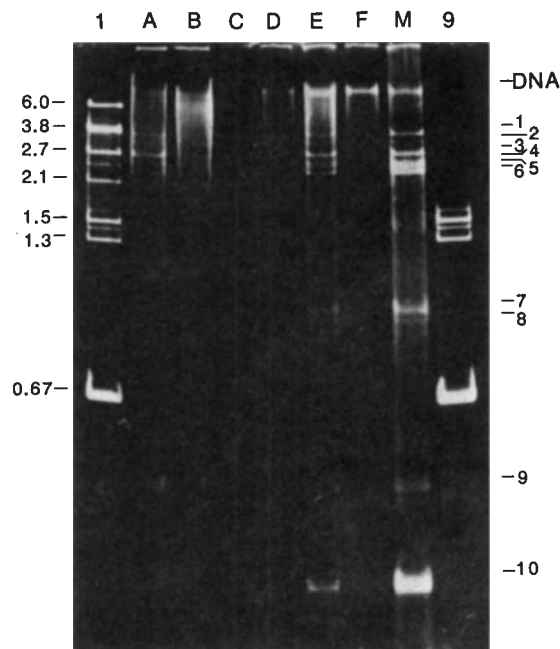


Fig. 2 Distribution of dsRNA segments in subcellular fractions from isolate Log1/3-8d². Double-stranded RNA prepared from each fraction was resuspended in the same volume of buffer, so that the intensities of the dsRNA bands reflect their relative amounts in each fraction.

Methods. Isolate Log1/3-8d² was cultured as described⁵ and the mycelium (10 g) was mixed with 15% (w/v) sucrose in 20 ml TE buffer (100 mM Tris-Cl, 0.2 mM EDTA, pH 7.5) and glass beads (20 g, Braun glasperl, 0.45–0.5 mm diameter) and shaken in a Braun MSK homogeniser at 0–5 °C until most of the hyphal compartments were disrupted. The homogenate was mixed with an equal volume of 1% (w/v) sucrose in TE buffer and the nuclei and cell wall debris (fraction A) were pelleted by centrifugation at 1,300g for 10 min at 4 °C. The supernatant was collected and centrifuged at 14,000g for 30 min at 4 °C to produce a high-speed supernatant (fraction B) and crude mitochondrial pellets which were resuspended in 20% (w/v) sucrose in TE buffer (4 ml). For purification of the mitochondria, step gradients were prepared each consisting of a bottom layer of 60% (w/v) sucrose in TE buffer (18 ml); a middle layer of 35% (w/v) sucrose in TD buffer (18 ml) and a top layer of resuspended mitochondrial pellets (2 ml), and centrifuged at 64,000g for 90 min at 4 °C in a Beckman SW28 rotor. The gradients were divided into the 20% sucrose layer (fraction C), the 35% sucrose layer (fraction D), the mitochondrial fraction which formed a dense brown band at the interface between the 35% and 60% sucrose layers, and pellets which formed at the bottom of the 60% sucrose layer (resuspended in 1 ml of TE buffer and designated fraction E). The mitochondrial fraction was diluted slowly with 3 vols of 10 mM Tris-Cl, 0.2 mM EDTA, pH 8.0 and centrifuged at 14,000g for 30 min at 4 °C. The pellet was resuspended in 50 mM Tris-Cl, 5 mM EDTA, 50 mM NaCl, pH 8.0 (1 ml) to give the purified mitochondria (fraction M) and a supernatant (fraction F). Double-stranded RNA was extracted from each fraction and purified by cellulose chromatography as described⁹, precipitated with ethanol, resuspended in TE buffer and analysed by polyacrylamide gel electrophoresis⁵. Double-stranded RNA segments were detected by ethidium bromide staining and ultraviolet fluorescence¹⁹. Sizes (kb) of marker viral dsRNA segments from *Aspergillus foetidus*^{21–23} (lane 1) and *Penicillium stoloniferum*¹⁸ (lanes 1 and 9), and segment numbers of the dsRNAs of Log1/3-8d² are shown on the left and right sides of the gel respectively.

lysed by polyacrylamide gel electrophoresis (Fig. 2). It is clear that, with the exception of segment 9, the mitochondrial fraction contains more of the dsRNA segments than any other fraction. The dsRNA segments could still be detected after incubation of purified mitochondria with RNase using conditions known to degrade dsRNA. Furthermore when *Aspergillus foetidus* viral

Table 1 Origins and properties of healthy and d²-infected isolates of *O. ulmi*

Isolate	Origin	Healthy (H) or diseased (D)	Tolerant (T) or sensitive (S) to MBC fungicide	Segments of dsRNA
Log1/3-8d ²	England	D	T	1 to 10
Con18	Conidial isolate from Log1/3-8d ²	H	T	2, 6
Con6d ²	Conidial isolate from Log1/3-8d ²	D	T	1 to 10
H351	Belgium	H	S	None
H351d ²	Log1/3-8d ² + H351	D	S	1 to 10
Asc201	Ascospore isolate from Log1/3-8d ² (♀) × H351 (♂)	H	S	None
Asc201d ²	Log1/3-8d ² + Asc201	D	S	1 to 10
Con18d ²	H351d ² + Con18	D	T	1 to 10

All isolates belong to the North American (NAN) race of the aggressive sub-group. Note that + denotes transfer of d²-factor from donor to recipient by hyphal anastomosis; × denotes a sexual cross.

dsRNA (Fig. 2) was added to purified mitochondria and the mixture incubated with RNase A, the *A. foetidus*, but not the *O. ulmi*, dsRNA segments were completely digested (not shown). Hence the Log1/3-8d² dsRNA segments, including the proportion of segment 9 found in the mitochondrial fraction, were protected from RNase digestion and had not fortuitously become bound to the outside of the mitochondria during purification. The presence of dsRNA segments, particularly segment 9, in other fractions could indicate that they have more than one subcellular location, or could result from release of dsRNA segments from mitochondria during purification.

In most fungi which have been examined, dsRNA segments are enclosed by a coat protein in isometric virus particles¹⁰. Exceptions are the dsRNA segments in hypovirulent strains of the chestnut blight fungus *Endothia (Cryphonectria) parasitica*, which have been shown to be located in lipid-rich vesicles¹¹, and the dsRNA formed by hybridization of transcripts of both strands of the mitochondrial genome in *Saccharomyces cerevisiae*¹². Electron microscope examination of subcellular fractions of *O. ulmi* isolate Log1/3-8d² failed to detect either isometric particles or vesicles (H.J.R., R. D. Woods, K.W.B. & C.M.B., unpublished data) and little dsRNA was detected in the subcellular fraction (Fig. 2, fraction B) in which, from their sedimentation coefficients^{10,11}, such particles or vesicles would

have been found if present. Furthermore using dot hybridization of mtDNA with a ³²P-labelled denatured dsRNA probe we could not detect copies of Log1/3-8d² dsRNA segments in the *O. ulmi* mitochondrial genome (unpublished results). Hence the *O. ulmi* dsRNA segments seem to be members of a class of autonomously replicating genetic elements different from any described hitherto in the fungi.

The close correlation of dsRNA segments 4, 7 and 10 with the disease phenotype⁵ suggests that one or more of these segments, either directly or by means of proteins they encode, may be responsible for the deficiency of cytochrome *aa*₃ in d²-infected isolates. This seems to be the first report of fungal mitochondria as the targets for deleterious effects associated with dsRNA segments, although dsRNA has recently been reported in the mitochondria of maize plants showing cytoplasmic male sterility¹³. The results raise interesting questions about the possibility of transport of dsRNA into the mitochondria and the site of replication and expression of the dsRNA segments, given the differences in the genetic codes used by mRNA in the mitochondria and cytosol (for review see ref. 8). Answers to these questions and elucidation of the mechanisms which cause reduction of cytochrome oxidase *aa*₃ in *O. ulmi* may help in the design of genetically manipulated d-factors suitable for the biological control of Dutch elm disease.

We thank J. M. Palmer and R. K. Poole for advice on cytochrome determinations, J. M. Palmer for the use of the DWL spectrophotometer, and the SERC for the award of a CASE studentship to H.J.R.

Table 2 Cytochrome ratios in mitochondria isolated from healthy and d²-infected isolates of *O. ulmi*

Isolate	Cytochrome ratios		
	<i>aa</i> ₃ /b	<i>aa</i> ₃ /(c + c ₁)	b/(c + c ₁)
Con18	0.79	1.3	1.6
H351	0.76	1.1	1.5
Asc201	0.87	1.2	1.4
Log1/3-8d ²	0.26	0.26	1.0
Con18d ²	0.13	0.14	1.1
H351d ²	0.06	0.07	1.1
Asc201d ²	0.10	0.09	0.8
Con6d ²	—*	—*	2.4

Isolates were cultured as described previously⁵. Mycelia were either disrupted mechanically or converted to protoplasts¹⁴ before isolation of mitochondria as described in Fig. 2. Cytochromes *aa*₃, b and c + c₁ were estimated from reduced minus oxidized difference spectra^{15,16}, scanned from 400 nm to 650 nm using a DWL split beam spectrophotometer (American Instrument), using the following wavelength intervals and absorption coefficients (mM cm⁻¹) (refs 16 and 17): 609 nm–630 nm (*aa*₃, 16; b, 0; c, -0.08); 560 nm–577 nm (*aa*₃, -0.326; b, 22; c, 0); 550 nm to 540 nm (*aa*₃, 0.63; b, -3.12; c + c₁, 19.1). Cytochrome ratios of mitochondria prepared from mechanically disrupted mycelium, or from protoplasts, of the same isolate differed by less than 10%.

* Cytochrome *aa*₃ was barely detectable in this isolate.

Received 6 July; accepted 25 August 1987.

- Brasier, C. M. *Adv. Pl. Path.* **5**, 53–118 (1986).
- Brasier, C. M. *Nature* **305**, 220–223 (1983).
- Brasier, C. M. in *Fungal Virology* (ed. Buck, K. W.) 177–208 (CRC, Boca Raton, Florida, 1986).
- Webber, J. F. *Pl. Path.* (in the press).
- Rogers, H. J., Buck, K. W. & Brasier, C. M. *Pl. Path.* **35**, 277–287 (1986).
- Brasier, C. M. & Gibbs, J. N. *Ann. appl. Biol.* **80**, 231–235 (1975).
- Sherald, J. L. & Sisler, H. D. *Pl. Cell Physiol.* **13**, 1039–1052 (1972).
- Bockelman, B., Osiewicz, H. D., Schmidt, F. R. & Schulte, E. in *Fungal Virology* (ed. Buck, K. W.) 237–283 (CRC, Boca Raton, Florida, 1986).
- Morris, T. J. & Dodds, J. A. *Phytopathology* **74**, 854–858 (1979).
- Buck, K. W. in *Fungal Virology* (ed. Buck, K. W.) 1–84 (CRC, Boca Raton, Florida, 1986).
- Hansen, D. R., Van Alfen, N. K., Gilles, K. & Powell, W. A. *J. gen. Virol.* **66**, 2605–2614 (1985).
- Beilharz, M. W., Cobon, G. S. & Nagley, P. *Nucleic Acids Res.* **10**, 1051–1069 (1982).
- Finnigan, P. M. & Brown, G. G. *Proc. natn. Acad. Sci. U.S.A.* **83**, 5175–5179 (1986).
- Marriott, A. C., Archer, S. A. & Buck, K. W. *J. gen. Microbiol.* **130**, 3001–3008 (1984).
- Williams, J. N. *Arch. Biochem. Biophys.* **107**, 537–543 (1964).
- Jones, C. W. & Poole, R. K. *Meth. Microbiol.* **18**, 285–328 (1985).
- Rifkin, M. R. & Luck, D. J. L. *Proc. natn. Acad. Sci. U.S.A.* **68**, 287–290 (1971).
- Bozarth, R. F., Wood, H. A. & Mandelbrot, A. *Virology* **45**, 516–523 (1971).
- Maniatis, T., Fritsch, E. F. & Sambrook, J. *Molecular Cloning, a Laboratory Manual* (Cold Spring Harbor Laboratory, New York, 1982).
- Sanger, F., Coulson, A. R., Hong, G. F., Hill, D. F. & Petersen, G. B. *J. molec. Biol.* **162**, 729–773 (1982).
- Ratti, G. & Buck, K. W. *J. gen. Virol.* **14**, 165–175 (1972).
- Buck, K. W. & Ratti, G. *J. gen. Virol.* **27**, 211–224 (1975).
- Buck, K. W. & Ratti, G. *J. gen. Virol.* **37**, 215–219 (1977).

Stabilization of charges on isolated ionic groups sequestered in proteins by polarized peptide units

Florante A. Quiocho, John S. Sack & Nand K. Vyas

Howard Hughes Medical Institute and Departments of Biochemistry and Physiology and Molecular Biophysics, Baylor College of Medicine, Houston, Texas 77030, USA

Electrostatic interactions are of considerable importance in protein structure and function, and in a variety of cellular and biochemical processes¹⁻³. Here we report three similar findings from highly refined atomic structures of periplasmic binding proteins. Hydrogen bonds, acting primarily through backbone peptide units, are mainly responsible for the involvement of the positively charged arginine 151 residue in the ligand site of the arabinose-binding protein, for the association between the sulphate-binding protein and the completely buried sulphate dianion, and for the formation of the complex of the leucine/isoleucine/valine-binding protein with the leucine zwitterion. We propose a general mechanism in which the isolated charges on the various buried, desolvated ionic groups are stabilized by the polarized peptide units. This mechanism also has broad application to processes requiring binding of uncompensated ions and charged ligands and stabilization of enzyme reaction charged intermediates, as well as activation of catalytic residues.

L-Arabinose-, sulphate- and leucine/isoleucine/valine binding proteins are members of a large group of periplasmic proteins (collectively 'binding proteins') which serve as initial high-affinity receptors of bacterial active transport systems for a variety of nutrients and of sugar chemotaxis. Well-refined crystal structures reveal binding protein molecules that are remarkably similar in spite of the small degree of amino-acid sequence similarity⁴⁻⁷. All are ellipsoid and are composed of two similar globular domains connected by three separate segments. The arrangement of the two domains gives rise to the ellipsoidal shape and the deep cleft between the two domains, where the substrate-binding site is.

In the L-arabinose-binding protein, Arg 151 is one of six polar residues forming a total of ten strong hydrogen bonds with the bound L-arabinose substrate⁴. The refined 1.7-Å structure (*R*-factor = 14%) reveals that the arabinose substrate and many of the essential residues are buried in the binding site cleft between the two domains (Fig. 1). The positively charged arginine residue is completely inaccessible to the bulk solvent (as deduced by the method of Lee and Richards⁸), and all five hydrogen-bond donor groups of the guanidinium are used in hydrogen-bonding to the sugar and other residues—notably to backbone peptide CO groups (see Fig. 2).

Figure 3 shows the atomic details of the hydrogen-bonding between the sulphate-binding protein and the sequestered sulphate dianion, as recently determined at 1.7-Å resolution (*R*-factor = 16%). The sulphate is tightly held in place by seven hydrogen bonds; the mean value of the hydrogen-bond distances is 2.77 (±0.07) Å and that of the hydrogen-bond angles 156 (±12)°. Five of the hydrogen bonds are formed with main chain peptide NH groups.

Whereas the structures of the L-arabinose- and sulphate-binding proteins were determined with bound endogenous substrates, the 2.4-Å structure of the native Leu/Ile/Val-binding protein is of the unliganded form (unpublished data, see also refs 7 and 9). Recent refinement of the structure of the native binding protein crystal soaked in a leucine solution at 2.8-Å resolution (*R*-factor = 18%) indicates that the substrate is bound to one domain mainly by six hydrogen bonds, five involving peptide units (Fig. 4). The bound leucine is completely dehydrated and is only partially accessible (~15%) to the bulk solvent.

In the three interactions described above, we have not only assessed solvent accessibility but also concluded that the sulphate, the leucine, and Arg 151 residue are in the ionic or fully charged forms (Figs 2-4). The large number of strong hydrogen bonds formed in these interactions reduce not only any need for aqueous solvation but also the cost of burying the charges. Moreover, as the hydrogen bonds have fewer flexible degrees of freedom than water, they offer a more stable solvation shell for the substrates. Numerous van der Waals' contacts are also formed in the protein-substrate complexes.

Note that the peptide units associated with each of the ionic groups are additionally coupled to hydrogen-bond arrays which

Fig. 1 Stereoscopic view of the C_α backbone trace of the refined 1.7-Å structure of L-arabinose-binding protein. The ligand-binding-site residue Arg 151 and the bound L-arabinose (Ara) (drawn as ball-and-stick model) are located in the cleft formed between the two domains. N, amino terminal end; C, carboxyl end. Numbers are given for every 20 C_α atoms. Note that Arg 151 is surrounded by three N termini of three parallel helices.

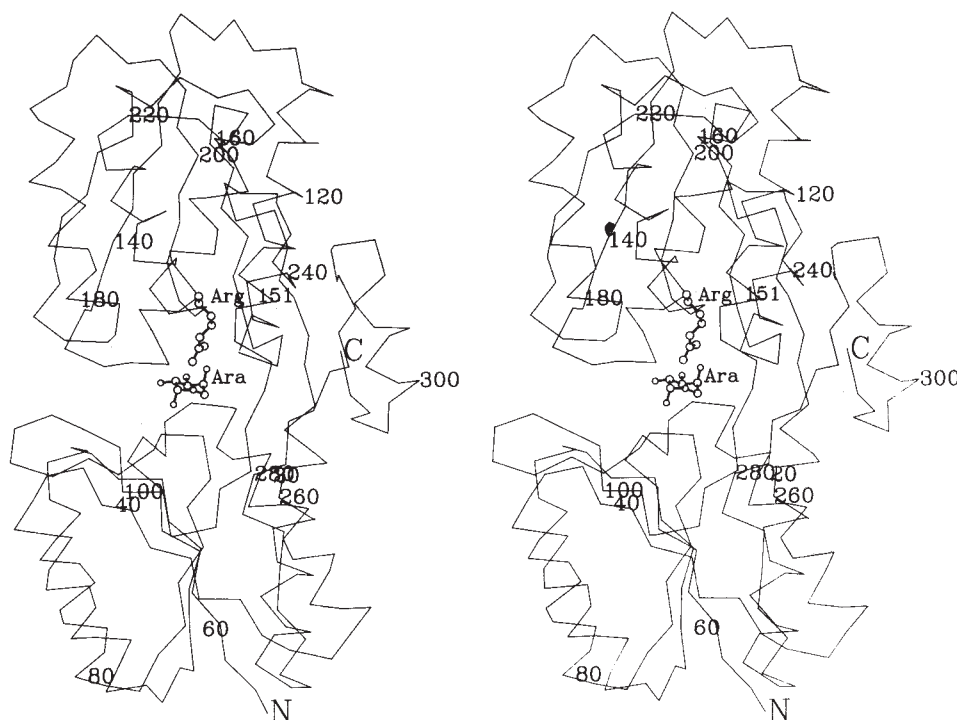


Fig. 2 The hydrogen-bonding interaction of Arg 151 of the L-arabinose-binding protein. Double circles, the non-hydrogen atoms of the Arg 151. The $N\epsilon H$ donates a hydrogen bond to the peptide CO of residue 148. The $N\eta_1 H_2$ donates a hydrogen bond to $O\gamma$ of the Thr 147 and another hydrogen bond to the O-5 ring oxygen of the bound L-arabinose. The $N\eta_2 H_2$ donates two hydrogen bonds—one to the peptide CO of residue 204 and the other to the O-4 oxygen of L-arabinose. Arg 151 is further linked to two hydrogen arrays: the hydrogen-bonded peptide units 147–148 and 145–146 are part of an array starting at $N\epsilon H$ of Arg 151 and terminating at a water molecule; the peptide unit 204–205 has its CO group accepting an H bond from the Arg 151 $N\eta_2 H_2$ and its NH group forming a bifurcated H bond with Asn 205 $O\delta_1$ and the peptide N of Met 204. That each of the five donor groups of the side chain is favourably paired with one hydrogen-bond acceptor group indicates a positively charged arginine residue. Further evidence for a guanidinium group is as follows. Crystals of the L-arabinose-binding protein used in the structure determination were kept in mother liquor of pH 6.5 (ref. 21), about seven pH units below the normal pK_a of a guanidinium group. Also the protein in solution exhibits a broad pH range of maximal sugar-binding activity from pH 5 to 9 (ref. 22). The polar environment of the arginine residue is in line with a more normal pK_a .

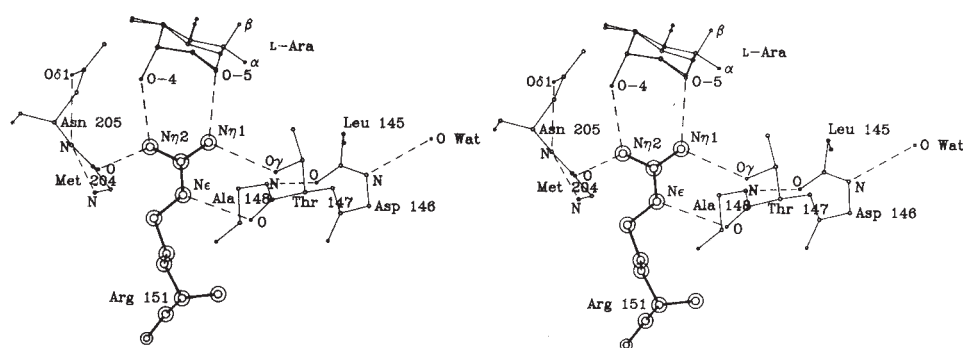
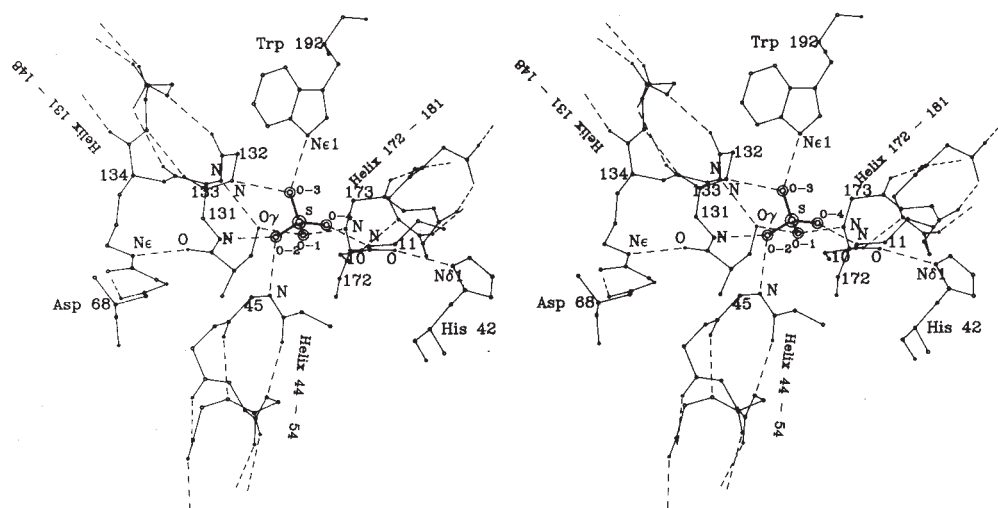


Fig. 3 The structure of the complex of the sulphate-binding protein with sulphate substrate as determined by crystallographic refinement at 1.7 Å resolution. The sulphate molecule is distinguished by double circle atoms. Only the first two turns of the three helices (positions 44–54, 131–148, 172–181) close to the sulphate are shown. The seven hydrogen bonds and five hydrogen-bond arrays associated with the sulphate oxygens are as follows: O-1 of the sulphate accepts two hydrogen bonds—one from the peptide NH of Ala 173 and the other from the $O\gamma H$ of Ser 130. The CO of peptide unit between residues 172 and 173 initiates a continuous neutral array of alternating peptide units and H bonds within an α -helix. This left-handed 'zig-zag' array of hydrogen-bonded peptide units has the general form: $CO_i \cdots NH_{i+4}$, where $i = m + 3(n - 1)$; m is the residue bearing the first peptide CO of the helix, and $n = 1, 2, 3 \dots$ indicates the extent of the peptide units linked by hydrogen bond. The array in this case has $m = 172$ and $n = 1$ to 3. O-2 is the recipient of two NH hydrogen bonds from the peptide units of Gly 131, and Ser 45 and is linked to two arrays: a neutral array within a helix with $m = 44$ and $n = 1$ –3 and an ionic array formed by the hydrogen-bonding of the CO group of the peptide unit 130–131 with the $N\epsilon H$ of the guanidinium of Arg 134, which is in turn salt-linked to Asp 68. O-3 accepts two H bonds with NH groups—one from the peptide bond between residues 131 and 132 and the other from the side chain of Trp 192. The CO group of the same peptide unit in turn initiates a neutral array within a helix with $m = 131$ and $n = 1$ –5. Finally, the peptide bond between residues 10–11 has its NH group donating the one hydrogen bond to O-4 of the sulphate and its CO group forming an ionic array with the $N\delta H$ of the imidazole ring of His 42 (assumed to be protonated at pH 4.3). Note that Ser 130 is also involved in an H-bonded array: the $O\gamma H$ donates an H bond to O-1 of the sulphate and accepts an H bond from the peptide NH of residue 133, which is in turn part of an array in a helix with $m = 132$ and $n = 1$ to 5. The evidence for a bound, fully ionized sulphate dianion is as follows. Although the structure of the binding-protein-sulphate complex was determined at pH 4.5, only 2.5 pH units above the pK_2 of sulphuric acid, it is unlikely that HSO_4^- exists as a stable species because there is no group in the binding site to accept a hydrogen bond and thereby stabilize the 'active' hydrogen. The lack of hydrogen-bond acceptors near the sulphate and a pK_1 of about -3 argues even more strongly against a sulphuric acid being bound to the sulphate-binding protein. Moreover, sulphate-binding activity is unaffected by pH (from pH 4 to 9)²³.

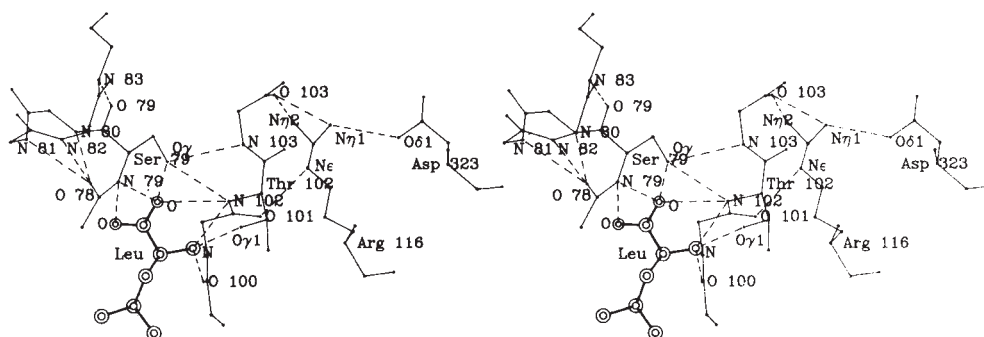


lead to the bulk solvent. There are two types of arrays: 'neutral arrays' are those formed with other main-chain peptide bonds, and 'ionic arrays' are those in which the peptide units are additionally hydrogen-bonded to charged side-chain residues. The three arrays (two neutral and one ionic) associated with the Arg 151 in L-arabinose-binding protein are shown in Fig. 2. Five arrays (three neutral and two ionic) are coupled to the sulphate (Fig. 3) and four arrays (three neutral and one ionic) to the bound leucine (Fig. 4). As three of the peptide NH groups

hydrogen-bonded to the sulphate are the first peptide units of α -helices, the sulphate is close to, but not centred on, the N-terminus of three helices (Fig. 3). Although the three helices are amphipathic, the left-handed arrays in these helices (formed by alternating main-chain peptide units and hydrogen bonds) coupled to the sulphate are on the side of each helix exposed to the bulk solvent.

In summary, despite differences in the nature and extent of the charges on the sulphate and leucine substrates and the

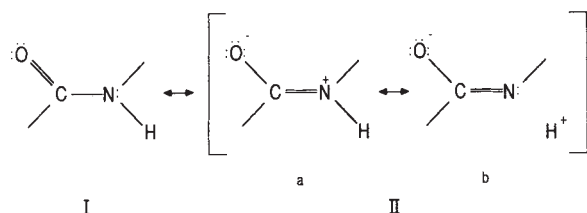
Fig. 4 The hydrogen-bonding of the L-leucine substrate to a major portion of the binding site of the Leu/Ile/Val-binding protein as determined at 2.8-Å resolution. Double circles, the non-hydrogen atoms of the leucine substrate; N, the α -ammonium group; O, the two oxygens of the α -carboxylate group. The α -ammonium group donates three hydrogen bonds, one each to the peptide CO



oxygen of residue 100 and O γ and peptide N of Thr 102. One of the oxygens of α -carboxylate group of the Leu substrate is the recipient of hydrogen bonds from the peptide NH of residue 79, O γ H of Ser 79, and peptide NH of residue 102. The other carboxylate oxygen is hydrogen-bonded with peptide NH of residue 79. The hydrogen-bond arrays mediated by the main-chain peptide bonds associated with leucine substrate are also shown. The NH group of residue 101 and hydroxyl group of Thr 102 are directed towards the wide-open cleft. Because the amino-acid substrates of the Leu/Ile/Val-binding protein contain weakly ionizing groups, the state of ionization of the bound leucine at about pH 6, the pH at which the structure was determined, is not fully certain. Nevertheless, the insensitivity of ligand-binding activity to pH (from pH 4 to 9)²⁴, and the nature of interactions shown, indicate a zwitterionic leucine substrate.

essential residue Arg 151, three key common factors govern the interactions of each of these buried ionic groups with the specific binding protein. First, hydrogen bonds are involved in the interactions; remarkably, there are no counter-charged residues or ions and water molecules. Second, the hydrogen bonds are formed chiefly with main-chain peptide units: the peptide CO groups being associated with the positively charged Arg residue, the peptide NH moieties with the sulphate dianion, and both peptide groups with the leucine zwitterion. And third, the peptide units associated with these charged groups are in turn coupled to hydrogen-bond arrays. The net effect of these three features are highly polarized peptide bonds. Interestingly, the three features may be more common in a variety of electrostatic interactions than previously recognized.

The three key factors above are the primary basis of the proposed mechanism for stabilization of isolated charges on the different ionic groups. The mechanism also rests upon two features of the peptide bond: it is composed mainly of two canonical resonance structures—a neutral [I] and ionic [IIa] species—in almost equal proportions¹⁰, and the labile electronic structure of the peptide bond can be modulated by a variety of interactions such as hydrogen-bonding and chelation (refs 11–13).



Thus, the mechanism simply considers the hydrogen bonds—particularly those involving the highly polarized peptide units—to be effective in stabilizing the isolated charges on the Arg residue, the sulphate and the leucine. The ionic tautomeric forms [II] of these peptide units ought to be stabilized, thus providing enough highly localized partial counter-charges to neutralize the different ionic groups. The arrays of hydrogen bonds coupled to the sequestered ionic groups could further disperse the buried charges.

The presence of several helices in the binding-protein structures^{4,5,7,9}, together with the concept that the alignment of peptide dipoles parallel to the helix axis may give rise to a macrodipole with partial charges^{12,14–15}, suggests that helix macrodipoles could provide some means of stabilizing the charges on various ionic groups. (The N-terminus of the helix

is the positive end whereas the C-terminus is the negative end.) Each of the two lobes of the binding proteins contains at least four parallel helices oriented in such a way that the helix N-termini are pointing at the ligand-binding-site cleft formed between the two lobes, and the C-termini are at opposite exposed ends of the elongated protein (for example, see Fig. 1).

Helix macrodipoles could assist in the binding of sulphate, although they have no effect on Arg 151 and the bound leucine. In fact, the possible involvement of helix termini is not a common feature of the electrostatic interactions involving all three charged groups. There is a total of nine α -helices in the arabinose-binding protein, but not a single helix C-terminus is near Arg 151 (Fig. 1). In fact, as shown in Fig. 1, the N-termini of eight helices converge at the cleft where the uncharged arabinose substrate and the positively charged arginine residue are situated. Although there are seven helices in the domain of the Leu/Ile/Val-binding protein where the substrate is bound⁷, the ammonium and carboxylate groups of the bound leucine are nowhere near a helix terminus. Three helix N-termini are close to, but not centred on, the sulphate bound in the sulphate-binding protein (Fig. 3). These three and the five other helix N-termini that are directed at the binding-site cleft⁵ could assist in long-range attraction of the sulphate. On the other hand, in the final complex formation, the more highly directional hydrogen bonds and peptide dipoles are crucial, not only in conferring specificity and affinity on the binding site and ensuring correctness of fit for the substrate, but also in stabilizing the buried charges. Hydrogen bonds are also important in the binding of sugars⁴ and amino acids (Fig. 4), but there is no indication that helices might be involved in either long- or short-range binding of neutral or zwitterionic substrates. Indeed, the fact that ligands of the entire family of periplasmic binding proteins, though often totally different (for example mono- and oligo-saccharides, inorganic ions, and every type of amino acid), are bound with similar affinity ($K_d \approx 1 \mu\text{M}$) precludes the influence of helix macrodipoles.

In the context of our proposal that polarized, highly directional peptide units are dominant in the specific binding of uncompensated charged ligands, we see two important functions of α -helices. First, the first turn of a helix and the adjoining loop serve as a source of concentrated, radially oriented hydrogen-bonding and charge-stabilizing groups, most of which are main-chain peptide units, and second, the helix contains three well-ordered hydrogen-bond arrays for charge dispersion. (Note that the loop and/or the helix amino-terminus are almost always populated by Gly residues as well as Ser or Thr residues. For example (Fig. 3), the peptide segment Ser 130-Gly 131-Gly

132-Ala 133, which forms part of a loop and the first turn of a helix, provides a complementary three-point docking site for O-1, O-2 and O-3 of the sulphate, respectively, and also provides three arrays, each initiated by the CO groups of the peptide units in the segment. We have indicated previously⁵ the extensive charge-stabilizing capacity of a six-residue peptide segment making up a loop and the first turn of a helix in the structure of the flavodoxin-FMN complex: this segment alone deploys eight of the ten hydrogen bonds to the FMN phosphate group located in a crevice close to the surface of the protein¹⁶. Likewise, peptide units of helices and adjoining loops clearly shown to be hydrogen-bonded to uncompensated phosphate groups of other coenzymes and of ligands such as DNA and sugars, can be considered in the context of our proposal. (The structure coordinates of flavodoxin and other proteins which were used in our analysis are available from the Protein Data Bank¹⁷.)

There is a component in the binding protein sulphate complex which strongly suggests a mechanism for activation of an essential residue by a polarized peptide unit. The simultaneous involvement of the hydroxyl of Ser 130 as a hydrogen-bond donor to O-1 of the sulphate and acceptor from the NH of residue 133, which is in turn linked by the CO group of residue 132 to an array within a helix (Fig. 3), could promote the sharing of the hydroxyl proton with O-1 of the sulphate, yielding a transient HSO_4^- . We observed two unusual features of this cooperative hydrogen-bonding in the 1.7-Å structure of the complex: The hydrogen-bond distance between the Ser hydroxyl oxygen and the sulphate oxygen (O-1) is somewhat short (2.69 Å), and the electron densities of the two oxygens are the only ones that are almost contiguous. These features are highly indicative of a charge-transfer complex.

Similarly, main-chain peptide-mediated hydrogen bonds could exert a more specific and significant influence in the activation or stabilization process in a variety of enzyme catalysed reactions. For instance, we note that chymotrypsin, elastase, subtilisin and many other related proteases have an important common feature in addition to the charge-relay system¹⁸: one or two main-chain NH groups (for example those of residues 192 and 193 in chymotrypsin) are responsible for stabilizing—by hydrogen bonds—the tetrahedral intermediate in the so-called 'oxanion hole'^{18,19}. (Interestingly, with the exception of papain and subtilisin, there are no helix N-termini close to the active site of proteolytic enzymes¹⁴.) There seems to be a similar oxanion hole in *p*-hydroxybenzoate hydroxylase for stabilizing the transient peroxide anion intermediate²⁰. Note that the peptide units associated with these 'holes' do not belong to helices. Rhodanese, we believe, also serves as a good example of the greater reliance on several peptide-unit-mediated hydrogen bonds in both the activation of an essential residue and the stabilization of a charged intermediate. In this enzyme, the negatively charged reactive cysteine residue and the persulphide anion intermediate make at least four hydrogen bonds with main-chain peptide NH groups²⁰. Besides the extensive use of main-chain peptide units, we have also observed that one or more of these units in the enzymes cited above are also coupled to hydrogen-bond arrays, further facilitating activation or stabilization. A highly advantageous feature in these two processes is the greater directionality of 'active site' peptide and hydrogen-bond dipoles in contrast to that of helix macrodipoles.

In conclusion, the results presented here from highly refined atomic structures are important not only because they constitute prime direct evidence that a diverse set of uncompensated charged ligands and residues can be sequestered in proteins but also because they provide the basis for understanding electrostatic interactions. Our results indicate that peptide unit-mediated hydrogen bonds are of paramount importance in stabilizing the buried, isolated charges and in electrostatic interactions of diverse nature. Moreover, the function of the peptide units does not necessarily rely on prior formation of secondary structure. Finally, the electrostatic interactions described here are not

greatly different from those of neutral ionophores (such as valinomycin), in that fixed, highly localized dipoles are also the most important factor in cation binding, specificity and charge stabilization. But, as clearly illustrated by our studies, proteins are more versatile, having in their arsenal a variety of means to neutralize all kinds of biologically important ions and charged ligands. Indeed, we have also recently discovered a novel calcium-binding site in the D-galactose-binding protein⁶.

This work was further supported by the NIH and Welch Foundation.

Received 31 July; accepted 18 August 1987.

1. Perutz, M. F. *Science* **201**, 1187–1191 (1978).
2. Matthews, J. B. *A Rev. Biophys. biophys. Chem.* **14**, 387–417 (1985).
3. Honing, B. H., Hubbel, W. L., & Flewelling, R. T. *A Rev. Biophys. biophys. Chem.* **15**, 163–193 (1986).
4. Quioco, F. A. & Vyas, N. K. *Nature* **310**, 381–386 (1984).
5. Pflugrath, J. W. & Quioco, F. A. *Nature* **314**, 257–260 (1985).
6. Vyas, N. K., Vyas, M. N. & Quioco, F. A. *Nature* **327**, 635–638 (1987).
7. Quioco, F. A. *et al.* in *Crystallography in Molecular Biology NATO Series* (eds Moras, D. *et al.*) 385–394 (Plenum, London, 1987).
8. Lee, B. K. & Richards, F. M. *J. molec. Biol.* **55**, 379–400 (1971).
9. Saper, M. A. & Quioco, F. A. *J. biol. Chem.* **258**, 11057–11062 (1983).
10. Pauling, L. *Nature of the Chemical Bond* 2nd edn (Cornell University Press, Ithaca, New York, 1960).
11. Llinas, M. & Klein, M. *J. Am. chem. Soc.* **97**, 4731–4737 (1975).
12. Wada, A. *Adv. Biophys.* **9**, 1–63 (1976).
13. Van Duijn, P. Th. & Thole, B. T. *Biopolymers* **21**, 1749–1761 (1982).
14. Hol, W. G. J., van Duijn, P. T. & Berendsen, H. J. C. *Nature* **273**, 443–446 (1978).
15. Hol, W. G. J. & Wierenga, R. K. in *X-Ray Crystallography and Drug Design* (eds Horn, A. S. & de Ranter, C. J.) 151–168 (Clarendon, Oxford, 1984).
16. Smith, W. W. *et al.* *J. molec. Biol.* **117**, 196–225 (1980).
17. Berstein, F. C. *et al.* *J. molec. Biol.* **112**, 535–542 (1977).
18. Kraut, J. A. *Rev. Biochem.* **46**, 331–358 (1977).
19. Robertus, J. D. *et al.* *Biochemistry* **11**, 4293–4293 (1972).
20. Wierenga, R. K. *Trans. Am. crystallogr. Ass.* **22**, 49–62 (1986).
21. Gilliland, G. L. & Quioco, F. A. *J. molec. Biol.* **146**, 341–362 (1981).
22. Parsons, R. G. & Hogg, R. W. *J. biol. Chem.* **249**, 3602–3607 (1974).
23. Pardee, A. B. *J. biol. Chem.* **241**, 5886–5892 (1966).
24. Penrose, W. R. *et al.* *J. biol. Chem.* **243**, 5921–5929 (1968).

Copies of articles from this publication are now available from the UMI Article Clearinghouse.

For more information about the Clearinghouse, please fill out and mail back the coupon below.

UMI Article Clearinghouse

Yes! I would like to know more about UMI Article Clearinghouse.

I am interested in electronic ordering through the following system(s):

- ☐ DIALOG/Dialorder ☐ ITT Dialcom
☐ OnTyme ☐ OCLC ILL Subsystem

☐ Other (please specify) _____

☐ I am interested in sending my order by mail.

☐ Please send me your current catalog and user instructions for the system(s) I checked above.

Name _____

Title _____

Institution/Company _____

Department _____

Address _____

City _____ State _____ Zip _____

Phone (_____) _____

Mail to: University Microfilms International
 300 North Zeeb Road, Box 91 Ann Arbor, MI 48106

Biosensors emerge from the laboratory

J. Briggs

Sensitive, stable and reliable biosensors are now available for the development of sophisticated bioanalytical measurement systems for applications in biotechnology.

THE expectation that biosensors will have a major impact on bioanalytical measurements has been with us for over a decade. Often high expectations have faded to scepticism as sensing schemes have proved unable to deal with the practical problems of working with real samples outside the research laboratory. The challenge and the magnitude of opportunity for biosensor applications in both medical diagnostics and the analysis of bioprocessing procedures is large. In bioprocessing, early biosensor systems will address measurement needs in the laboratories which support the development and production of biotechnology products.

There are many different types of biosensor, each operating on different physical principles, and each with its own repertoire of problems and potential applications. In the broadest sense, a biosensor consists of a bioactive layer of enzymes, antibodies or receptors, in intimate contact with a transducer. The transducer responds to biochemical reactions occurring in the bioactive layer by sending out an optoelectronic or purely electronic signal. If the biochemical reaction has been modulated by the presence of a particular analyte in a sample, then the output signals can be interpreted to provide information about the prevalence of that analyte in the sample.

Traditional analytical techniques probe a sample with a source such as a light beam, detect the response with a detector, and then convert the response into an electronic signal away from the sample. Rapid developments in semiconducting electronics have led to attempts to bring the electronic amplification step closer to the biochemical sample. This approach reduces the noise generated in the front-end of the system — the chemistry and first stages of electronics — where the specific signal is weak, and hence increases the signal-to-noise ratio.

The incorporation of semiconducting electronics into a biosensor allows the production of small and potentially inexpensive devices which could be used in applications such as *in vivo* medical diagnostics and on-line monitoring of bioprocesses. A large amount of effort has been devoted to research on ChemFETs (chemical field effect transistors)¹. A FET operates on the principle that the electrical field at the gate region of the transistor modulates the thickness of a current-carrying channel in the semiconductor, which connects the source and drain

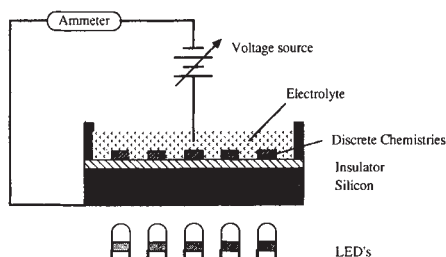


Fig. 1 The Molecular Devices biosensor. Courtesy H. McConnell, W. Parce and D. Hafeman. terminals. Small changes in the electrical field result in relatively large variations in the current, thereby amplifying the signal. In practice, a fluid is brought into contact with the gate region, where a biochemical reaction drives the electric field.

Unfortunately, there are several fundamental problems with these devices. Biochemical samples are nearly always rich in sodium ions, and sodium diffusion into silicon destroys the action of the transistor. The layer separating the gate from the semiconductor, therefore, must be thin, impervious to ion diffusion, and a very good electrical insulator. The biochemical fluid must also be isolated from the transistor, because the slightest amount of leakage of the electrolyte will short out the device. Though these problems surely will be solved, to date FET-based sensors have limited practicality.

Surface phenomena on small devices are currently being investigated as potential biosensors. The absorption of a monolayer, mediated by an antibody-antigen reaction, has been found to change appreciably the reflectivity of a surface², the absorption of an evanescent wave by an optical fibre^{3,4}, and the propagation of surface acoustical waves⁵. Surface phenomena techniques yield a direct response to a biochemical reaction involving the analyte molecule, without the need for a secondary enzyme or radioactive label. But discrimination of the analyte from nonspecific surface binding of other proteins and compounds in biochemical samples has been a major challenge for surface phenomena-based techniques.

Molecular Devices has developed a silicon-based biosensor with a proprietary insulating layer as the monolithic top surface (Fig. 1). The silicon, insulator and electrolyte, when combined, act as a capacitor. A closed circuit is achieved with the addition of an ammeter, an electrode, a means of varying the voltage, or bias, across the capacitor, and the attachment of a wire to the back of the silicon. To

make a measurement, the back side of the biosensor is illuminated from one of an array of light-emitting diodes (LEDs). When a given light is turned on, the photo-response in conjunction with the reverse biasing of the device causes the interface between the silicon and the insulating layer to charge up. With this transient response, an impulse of current flows in the external circuit. By modulating the light (typically at 10 kHz), an alternating current is generated, which is measured by the ammeter. No direct current passes through the sensor, and the magnitude of the photogenerated alternating current depends upon the bias potential across the biosensor. An enzyme reaction on the biosensor surface, at one of the discrete chemistry locations, shifts the surface potential and hence the current. The amount of change of the external potential is a direct measure of the enzyme activity, required to compensate for the change in surface potential, or number of enzyme molecules when the substrate is in excess. The direct conversion of an enzyme reaction to an electrical response leads to highly sensitive detection of analyte. Only 300,000 enzyme molecules with high turnover are required to produce a measurable response in one minute, over the background drifts of the sensor⁶; orders of magnitude lower than what can be detected colorimetrically.

The photogenerated, alternating current only flows vertically through the device at the point where the light is on and modulated, but there are two primary mechanisms by which signals can spread horizontally on the biosensor. First, the photocarriers can diffuse laterally in the bulk silicon, but only over a distance comparable to the thickness of the wafer.

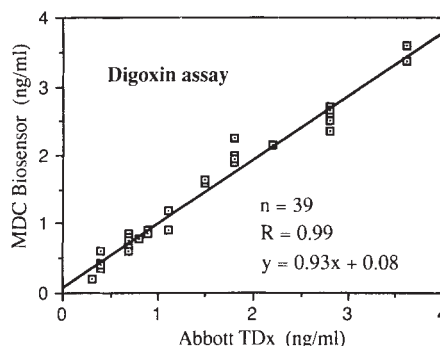


Fig. 2 Comparative results of an assay for the drug digoxin. Thirty-nine patient serum samples were run with no sample pretreatment, and were read with the Molecular Devices biosensor. Courtesy R. Zuk and R. Armenta. The reference method was the Abbott TDx system.

Second, the product of the chemical reaction at a discrete site may diffuse, but only over the time it takes to complete a measurement, which is typically one minute. In actual practice, both of these mechanisms result in spreading of only a few millimetres. Therefore, multiple localized measurements, corresponding to discrete chemistries, can be made with a monolithic sensor and one wire connection. This allows the direct incorporation of controls and calibrators for analytes that are difficult to measure.

The feasibility of this biosensor assay has been demonstrated using a broad spectrum of analytes, including direct assays for enzymes, immunoassays for molecules and microorganisms, and DNA probe assays. Figure 2 compares the results of immunoassays for the drug digoxin in human serum. Readings obtained with the biosensor compare well with the reference method based on fluorescence-polarization. In this case the enzyme-substrate pair, urease-urea, directly changes the local pH at the sensor surface, which in turn changes the surface potential. Measurements can also be made in the redox mode; hence, this biosensor can be linked to essentially any enzyme-mediated assay procedure.

The Molecular Devices biosensor is designed to overcome several practical difficulties. Because there is no structure in the silicon to preserve, the insulating layer is processed to be very stable and impervious to ion diffusion. Wafers have been stored in saline for over one year, and individual sensors have been used for many hundreds of determinations, with no deterioration in performance. The monolithic top surface makes sealing of fluids straightforward and facilitates the coupling of the biosensor to a wide variety of membranes and chemistries.

All emerging biosensor technologies face the challenge of going from feasibility demonstrations to successful first products. Early efforts should focus on the reduction to practice of biosensor technology, to provide a base for later specialized applications. Sophisticated biosensor-based systems to provide for laboratory measurements of samples derived from a bioprocess should be developed before sensors for in-line monitoring of the bioreactor, that will initially provide quantitation of analytes such as pH and redox potential. □

Jonathan Briggs is at Molecular Devices Corporation, 3180 Porter Drive, Palo Alto, California 94304, USA. For more information, fill in reader service number 100.

1. Janata, J. *Solid State Chemical Sensors*. 65–118 (Academic, New York, 1985).
2. Giaever, I. *J. Immunol.* **110**, 1424 (1973).
3. Kronick, M.N. & Little, W.A. *J. Immunol. Meth.* **8**, 235–240 (1975).
4. Sutherland, R.M. *et al. J. Immunol. Meth.* **74**, 253–265 (1984).
5. Roederer, J.E. & Bastiaans, G.J. *Analyt. Chem.* **55**, 2333–2336 (1983).
6. McConnell, H. *et al.*, in preparation.

Biotechnology bonanza

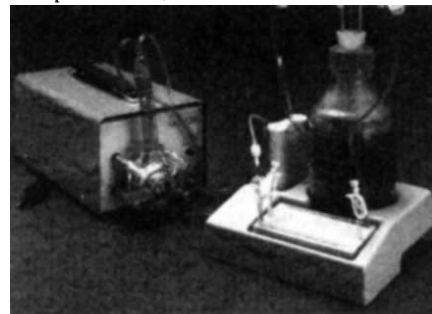
Engineering bacteria to produce proteins in vats and coaxing tissues to grow outside their native environs can be more art than science. This week offers the latest tools for taking the guess-work out of these tasks.

BOEHRINGER Mannheim Biochemicals offers enzymatic test kits for the determination of optimum media feed rates in bioreactors and tissue culture (Reader Service No. 101). The tests monitor quantitatively glucose consumption and lactic acid production — the two leading indicators of media exhaustion. The kits are currently used in Boehringer Mannheim's own R&D laboratories to monitor cell growth and control media feed rates. The kits include glucose-6-phosphate dehydrogenase, L-lactate dehydrogenase and associated buffers.

The American Type Culture Collection has six additions to its range of hybridomas that produce murine monoclonal antibodies against oncogene products (Reader Service No. 102). Myb 2-7.77 secretes IgG₁ that recognizes v-myb and chicken c-myb protein by immunoprecipitation and immunoblot analysis, while Myb 2-3.76 secretes IgG_{2b} that recognizes the same proteins. Human c-myc protein can be recognized by IgG₁ secreted by Myc CT 9-B7.3 through immunoprecipitation analysis, and by Western blot using IgG₁ from Myc 1-9E10.2. Myc CT 14-G4.3 secretes IgG₁ that recognizes human c-myc by immunoprecipitation and Western blot analysis. Viral and chicken c-myb proteins are recognized by immunoprecipitation and Western blot by IgG_{2b} secreted by Myb 2-37.63. These lines were deposited by J. M. Bishop and G. I. Evan, and cost \$45–72 (US) for a 1 ml frozen sample, containing 2.4×10^6 – 5.4×10^6 cells ml⁻¹.

Healthy environments

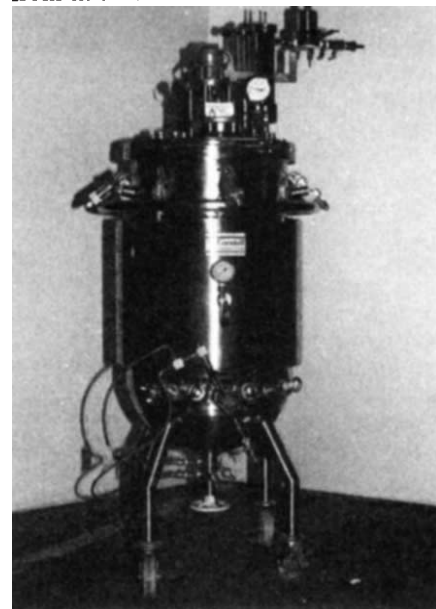
For continuous production of monoclonal antibody by cross-species hybridomas that cannot be grown in mice, Millipore offers its Dynacell membrane-based perfusion system (Reader Service No. 103). The cell module consists of nine series of alternating layers of cell culture compartments, media channels and mem-



More layers means better growth with Dynacell.

branes. All cells are evenly perfused with media, so the risk of creating static pockets in the cell compartments is minimized, allowing production to continue for several months. The approximately \$3,000 (US) Dynacell system is designed for optimal culture conditions: extremely high cell densities may be maintained, and cultures can gradually be weaned from serum. Millipore says the Dynacell system offers debris-free antibody production rates of up to 1.8 times greater than that of stirred culture systems.

A 140-litre production-scale tissue culture vessel is new from LH Fermentation (Reader Service No. 104). The 140VSTC is designed for the production of proteins from microcarrier-bound tissue cultures.

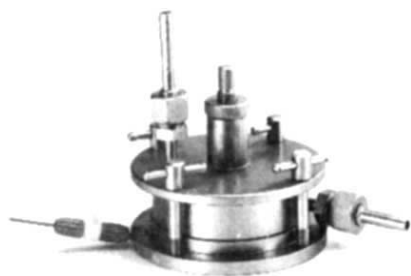


Supplied on casters for transportability.

It has a novel method for harvesting media: the stirrer shaft has a 70 µm rotating stainless steel mesh filter that screens out microcarriers, while allowing media to be drawn through a needle port on the top plate. The vessel has a hemispherical lower end, which provides ideal fluid flow conditions, and a four-bladed height-adjustable pitch blade impeller with a speed range of 20–200 r.p.m. The fermenter has a working volume of 100 litres, with a vessel aspect ratio of 1.9:1 and a liquid aspect ratio of 1.4:1.

Total control

Continuous sampling is a prerequisite for the efficient data monitoring and computer control of bioreactors. Braun has introduced its Biopem magnetically



Collects small samples for monitoring processes.

stirred filtration cell for the continuous aseptic **sampling of bioprocess fluids** by dynamic filtration (*Reader Service No. 105*). The £3,185 (UK) Biopem is autoclavable and *in situ* sterilizable, can be used in conjunction with standard laboratory analytical equipment, and can be disconnected and reconnected during fermentation. The magnetic stirrer can be adjusted to remove products which cause fouling of the 90 mm diameter membrane, which has an effective area of 45 cm².

The Questor **fermentation control system** from Extrel Corporation opens a window into industrial biotechnology processes to allow engineers to monitor production and waste accumulation during fermentation (*Reader Service No. 106*). Questor uses a quadrupole mass spectrometer filter and a DEC PDP 11 computer to sample and quantify the components of head-space gases. Extrel says the \$70,000–110,000 (US) system can analyse 16 components with molecular weights from 2 to 200 simultaneously, in concentrations from 10 parts per million to 100 per cent. Questor's software allows it to switch automatically from one process stream to another, to calibrate itself, and to log data. It also has the capability to scan through detected masses to identify contaminants. Questor can be linked to another computer or to trigger alarms.

Enprotech has a **computer-based bioprocessor** that can be used to control microbial and cell culture bioreactors produced by any major manufacturer (*Reader Service No. 107*). The system is equipped with colour graphics that indicate current control parameters and give a visual chart history of each run. Software guides the user to set-up for each run



Enprotech's system controls any bioreactor.

through step-by-step calibration procedures. Data can be retrieved for analysis with most popular statistical programs, and the system comes with battery backup and a sealed membrane keyboard to guard against spilled fluids. The complete system consists of software, hardware, bioreactor and probes, and costs \$20,000 (US). Components may be bought singly.

Biotechnology Computer Systems Ltd has introduced a PC-based version of its **Bio-i process management system**, called Bio-pc (*Reader Service No. 108*). Bio-pc is a single-user system that can control up to four bioreactors and associated on-line ancillary equipment at a time. It is IBM-compatible, and performs data monitoring, logging and feedback control in background mode, so that word-processing, spreadsheet, database management and other programs can be accessed at any time. The Bio-pc system retains the same format as the Bio-i system designed for large-scale fermentation.

Cultural growth

HL-1 is a bovine serum albumin-free, chemically defined **medium additive** formulated to support the growth of mouse hybridomas and lymphoid cells, made by Ventrex Laboratories and available from Tissue Culture Services Ltd (*Reader Service No. 109*). TCS says the low protein content of HL-1 — less than 30 µg protein ml⁻¹ when diluted 1:100 in the basal medium — improves yield in antibody production, and offers lower non-specific binding in screening assays. HL-1 can be stored for up to 60 days at room temperature, and is supplied as a 100 × concentrate in a 100 ml bottle, sufficient for 10 litres of media. TCS tests the HL-1 for sterility, pH, endotoxin, and protein concentration, and assesses its growth characteristics on three hybridoma cell lines before shipping.

Janssen has drastically reduced the cost of its **hybridoma growth factor**: 20 ml now costs £25 (UK) (*Reader Service No. 110*). Hybridoma growth factor is produced by human monocytes and has been shown to be important for the growth of B-cell hybridomas. It migrates on SDS-polyacrylamide gels as two 21 KD and 25 KD bands. The nearly 50 per cent reduction in price makes hybridoma growth factor considerably cheaper than crude IL-2 or human endothelial culture supernatant in the culture of hybridomas.

Bulk packs of liquid media are available under custom order from Gibco Laboratories (*Reader Service No. 111*). The media is supplied in 10-litre Stak Pak plastic packages that are lighter and easier to store than glass bottles. The Stak Pak can be suspended vertically or stacked on its side for serial connection, eliminating the



For laboratories with large media needs.

need for a holding tank. Stak Pak's feature sterile access ports for adding supplements and dispensing media aseptically.

Fermentech is offering to share its expertise in the development of **continuous fermentation processes** on a contract basis (*Reader Service No. 112*). Drawing upon its experience in developing hyaluronic acid and protein A for medical applications, the company is providing process development and contract fermentation research services. Fermentech says continuous fermentation offers advantages in increased physiological control of the organism, batch-to-batch consistency and lower production costs.

Downstream processes

From Waters, a division of Millipore, comes a family of reverse-phase **columns**



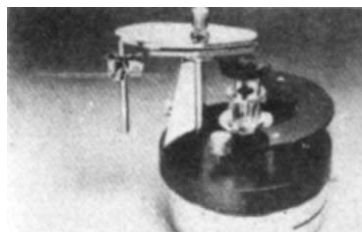
Move up in size as you scale up a process.

for scaling up preparative separations involving proteins, peptides, natural products and synthetic intermediates (*Reader Service No. 113*). The Delta-Pak columns come in C₁₈ and C₄ bonded phase chemistries, with pore sizes of 100 Å or 800 Å. Each of the five sizes, ranging from 3.9 mm × 30 cm to 50 mm × 30 cm, is available with the chemistry of choice. After the separation has been scaled up to the preparative column, the purity of collected fractions may be checked on a

ADVERTISEMENTS

SINGER INSTRUMENTS**MICROMANIPULATORS MICROFORGE
WORKING WITH YEAST?**

You will find it so easy
with a Singer Mk III
Micromanipulator



Singer Instrument Co. Ltd

Tel. (0984) 40226
Telex: 337492 Comcab

Reader
Service
No.50

**SENSORS AND
INSTRUMENTATION**

- ♦ Chemical microsensors for field or lab.
Sensing capabilities: CO₂, PO₂, PH₂,
pH, specific ions

FREE CATALOG

- ♦ New Clark-style Micro electrode
50-250 µm tip diameter
- ♦ Specialists in Mini/Micro electrodes
- ♦ Custom electrodes/instrumentation



Diamond General Corp.

3965 Research Park Drive
Ann Arbor, MI 48108 USA 313-747-9700

Reader Service No.13

ULTRA-AMERICA I./E. CO.

Agent for Laboratory Animal Centre,
Academy of Military Medical Science, P.R.
China, trading high quality animals with
quarantine health certificate. Please
contact:

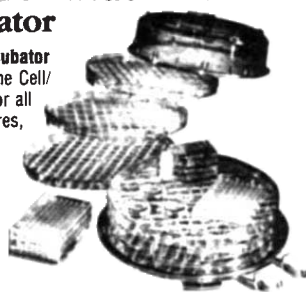
**Shop 18
Pelene Arcade
Yue Ko Street
Tin Wan
Hong Kong**

Reader
Service
No.21

**Versatile Tissue Culture
Incubator**

**Modular Incubator
Chamber (the Cell/
Sphere™)** for all
tissue cultures,
aerobic or
anaerobic.

**Low
Cost**



Eliminates evaporation loss and contamination. Easy
to use - flush with gas mixture, seal, place at appro-
priate temperature. Stackable, see-through. Quantity
discounts. billups-rothenberg, inc., pob 977, Del
Mar, CA USA 92014. Call 619-755-3309 or contact
your Flow Laboratories representative.

Reader Service No.174

higher efficiency 5µm column with the
same chemistry. The columns cost \$240-
4,900 (US) each, depending upon size.

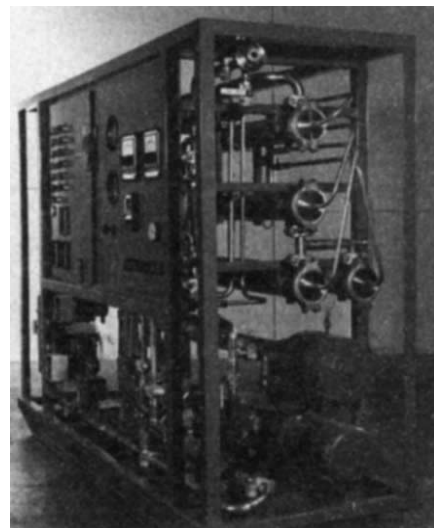
The **CELSEP cell separation system**
from Wescor uses the principle of velocity
sedimentation combined with chamber
reorientation to perform separations of
heterogeneous mammalian cell suspen-
sions (Reader Service No. 114). Wescor
says the system performs most separations
in 2-3 hours, with high reproducibility.
The system separates cells at unit gravity,
and approaches the maximum theoretical
resolution possible with velocity sediment-
ation. CELSEP costs \$1,995 (US).

Osmonics sells a system for **testing
membrane performance** in reverse osmo-



Wescor's brochure on the CELSEP.

sis and ultrafiltration applications up to
93°C. The OSMO process evaluation sys-
tem holds up to eight 4-in diameter
OSMO reverse osmosis or ultrafiltration
separators, or membrane elements
(Reader Service No. 115). It can evaluate
feed rates from 2 to 20 g.p.m. at pressures



For rent or for sale, the OSMO tests membrane
performance in a range of applications.

from 50 to 800 P.S.I.G. The OSMO is
microprocessor-controlled, and comes
with a pH controller and two chemical-
feed pumps for pretreatment. Osmonics
also offers the system for rent.

Production-scale columns have been
added to Sepragen's line of low-pressure,
radial-flow columns (Reader Service No.
116). The new Superflo columns can hold
10-100 litres of packing material, and
yield production rates up to 10 times
above those of conventional columns with
equivalent capacity, according to Sepra-
gen. The columns can be used with rela-
tively inexpensive packings, pumps and
associated hardware to reduce costs. The
columns cost \$8,500-19,000 (US) and
come in 40 × 32cm to 40 × 120 cm sizes.

Alarming situations

Omega Engineering has a dual relay
alarm module that adds alarm capability
to any control system loop (Reader Service
No. 117). The \$185 (US) PHA-51 can be
configured to set off an alarm when pH, or
any condition being monitored by an elec-
trical device, exceeds the desired range.

Reliable sources

Biotechnology patent activity in Japan can
be monitored with the *Japanese Biotech-
nology Patent Alert*, published in English
by the A. Fujiwara Scientific Liaison
Office (Reader Service No. 118). The
bulletin is issued 32 times per year, and
contains summaries of unexamined patent
applications within one month after they
are printed in the Japanese patent gazette.
The subscription price is \$650 (US), and
subscribers may also purchase full texts of
selected patents translated into English
for \$40 (US) per typed page, with a \$20
(US) handling charge per order.

The *Biotechnology Abstracts* published
by Derwent Publications Ltd are now
available online (Reader Service No. 119).
Access to the **abstracts** is through the
ORBIT Search Service, at \$100 per
connect hour. The abstracts are compiled
from over 1,000 scientific and technical
journals, conference proceedings, meet-
ing reports and patent literature. All the
data are searchable and printable, includ-
ing titles, indexing keywords, biblio-
graphic details and full abstract texts. The
abstracts cover all aspects of biotech-
nology from genetic manipulation
through biochemical engineering and
fermentation to downstream processing.
The database contains approximately
60,000 records, but Derwent says searches
normally take only a few minutes. □

*These notes are compiled by Carol Ezzell from
information provided by the manufacturers. To
obtain further details about these products, use
the reader service card bound inside the journal.
Prices quoted are sometimes nominal, and apply
only within the country indicated.*